

Foundations for a scaling-rotation geometry in the space of symmetric positive-definite matrices

David Groisser^{a,*}, Sungkyu Jung^b, Armin Schwartzman^c

^a*Department of Mathematics, University of Florida, Gainesville, FL 32611, USA*

^b*Department of Statistics, University of Pittsburgh, Pittsburgh, PA 15260, USA*

^c*Division of Biostatistics, University of California, San Diego, CA 92903, USA*

Abstract

We investigate a geometric structure on $\text{Sym}^+(p)$, the set of $p \times p$ symmetric positive-definite matrices, based on eigen-decomposition. Eigenstructure determines both a stratification of $\text{Sym}^+(p)$, defined by eigenvalue multiplicities, and fibers of the “eigen-composition” map $F : M(p) := SO(p) \times \text{Diag}^+(p) \rightarrow \text{Sym}^+(p)$. The fiber structure leads to the notions of *scaling-rotation distance* between $X, Y \in \text{Sym}^+(p)$, the distance in $M(p)$ between fibers $F^{-1}(X)$ and $F^{-1}(Y)$, and *minimal smooth scaling-rotation (MSSR) curves* [Jung et al. (2015)], images in $\text{Sym}^+(p)$ of minimal-length geodesics connecting two fibers. In this paper we study the geometry of the triple $(M(p), F, \text{Sym}^+(p))$, where $M(p)$ is given a suitable product Riemannian metric, and use this basic geometry to begin a systematic characterization and analysis of MSSR curves. This task raises basic questions: For which X, Y is there a *unique* MSSR curve from X to Y ? More generally, what is the set $\mathcal{M}(X, Y)$ of MSSR curves from X to Y ? This set is influenced by two ways that non-uniqueness can potentially occur, “Type I” and “Type II”. We translate the question of whether Type II nonuniqueness can occur into a question about the geometry of Grassmannians $G_m(\mathbf{R}^p)$, with m even, that we answer for $p \leq 4$ and $p \geq 11$. Our method of proof also yields an interesting half-angle formula concerning principal angles between subspaces of \mathbf{R}^p whose dimensions may or may not be equal. The general- p results concerning MSSR curves and scaling-rotation distance that we establish here underpin the $p = 3$ results in [8], where they enable us to find new explicit formulas for MSSR curves and scaling-rotation distance, and to identify $\mathcal{M}(X, Y)$, in all nontrivial cases.

Keywords: eigen-decomposition, stratified spaces, scaling-rotation distance, signed-permutation group, geometric structures on quotient spaces, principal angles, geometry of Grassmannians

2010 MSC: 53C99, 53C15, 57R15, 53C22, 51F25, 15A18, 58A35

[☆]This work was supported by NIH grant R21EB012177 and NSF grant DMS-1307178.

^{*}Corresponding author

Email addresses: groisser@ufl.edu (David Groisser), sungkyu@pitt.edu (Sungkyu Jung), armins@ucsd.edu (Armin Schwartzman)

1. Introduction

In this work, we investigate a geometric structure on $\text{Sym}^+(p)$, the set of $p \times p$ symmetric positive-definite (SPD) matrices, $p > 1$, and special curves that this structure gives rise to. Both the geometric structure and these special curves are built from eigen-decomposition of SPD matrices.

Let $\text{Diag}^+(p)$ denote the set of $p \times p$ diagonal matrices with positive diagonal entries. By an (orthonormal) *eigen-decomposition* of $X \in \text{Sym}^+(p)$ we will mean a pair $(U, D) \in SO(p) \times \text{Diag}^+(p)$ such that $X = UDU^{-1} = UDU^T$. The space of such decompositions,

$$M(p) := SO(p) \times \text{Diag}^+(p), \quad (1.1)$$

thus comes naturally equipped with a smooth surjective map $F : M \rightarrow \text{Sym}^+(p)$ defined by

$$F(U, D) = UDU^T. \quad (1.2)$$

For each $X \in \text{Sym}^+(p)$ we define the *fiber over X* to be the set $\mathcal{E}_X := F^{-1}(X)$.

Note, however, that $M(p)$ is not a fiber bundle over $\text{Sym}^+(p)$ with projection F ; the map F is not even a submersion. (Rather, the relation of $M(p)$ to $\text{Sym}^+(p)$ is reminiscent of the notion of *blow-up* in algebraic geometry: $M(p)$ can be viewed as a sort of blow-up of $\text{Sym}^+(p)$ along several subvarieties.) The natural action $SO(p) \times \text{Sym}^+(p) \rightarrow \text{Sym}^+(p)$, $(U, X) \mapsto UXU^T$, endows $\text{Sym}^+(p)$ with a stratification by orbit-type, and the derivative of F is nonsingular only on the pre-image of the “top” stratum. This stratification is identical to the stratification by “eigenvalue-multiplicity type”, in which the strata are labeled by partition of the integer p . Eigenvalue multiplicities also determine a more refined stratification of the space $M(p)$, in which the strata are labeled by partitions of the set $\{1, \dots, p\}$.

The fiber structure of $M(p)$ formalizes the notion of *minimal smooth scaling-rotation curves* [11], whose characterization and analysis are among the main goals of this paper. Motivated by applications to diffusion-tensor imaging, Schwartzman [13] introduced smooth scaling-rotation curves as a way of interpolating between SPD matrices in such a way that eigenvectors and eigenvalues both change at uniform speed. *Minimal smooth scaling-rotation curves* were defined in [11] as smooth curves whose length (as determined by an appropriate Riemannian metric on $M(p)$) minimizes the amount of scaling and rotation needed to transform an SPD matrix into another.

More precisely, each factor of $M(p)$ is a Lie group, and for our Riemannian metric g_M on $M(p)$ we take a product metric determined by choosing bi-invariant metrics $g_{SO}, g_{\mathcal{D}^+}$ on the factors $SO(p), \text{Diag}^+(p)$. We define *smooth scaling-rotation (SSR) curves* in $\text{Sym}^+(p)$ to be the projections to $\text{Sym}^+(p)$ of geodesics in $(M(p), g_M)$. In this scaling-rotation framework, the “distance” $d_{\mathcal{SR}}(X, Y)$ between any two matrices $X, Y \in \text{Sym}^+(p)$ is defined to be the

distance between the fibers \mathcal{E}_X and \mathcal{E}_Y (nonzero if $X \neq Y$ since each fiber is compact). The minimal-length geodesics connecting two fibers \mathcal{E}_X and \mathcal{E}_Y give rise to minimal smooth scaling-rotation curves (MSSR) curves, “efficient” smooth scaling-rotation curves that join X and Y .

The geometry of $M(p)$ with the Riemannian metric g_M is relatively simple; for example the squared-distance function d_M^2 is simply a sum of squares. Despite this simplicity “upstairs”, the problem of determining MSSR curves between arbitrary X, Y in the quotient space $\text{Sym}^+(p)$ is far from trivial, and the dependence on X and Y of the set $\mathcal{M}(X, Y)$ of such curves is quite intricate. Each fiber \mathcal{E}_X is a submanifold of $M(p)$, typically with many connected components, whose diffeomorphism-type (which we give very explicitly in Corollary 2.9) depends on the stratum of $\text{Sym}^+(p)$ in which X lies, and has positive dimension when X does not lie in the top stratum. While it is easy to write down an explicit formula for the distance between two arbitrary *points* of $M(p)$, even in Euclidean space there is no formula for distance between *submanifolds* of positive dimension, even when the submanifolds come equipped with explicit parametrizations or are level-sets of explicitly given functions. Understanding the fibers of F is thus one of the key ingredients in analyzing non-trivial features of the scaling-rotation framework. Among the nontrivial features is the fact, shown in [11], that $d_{\mathcal{SR}}$ is a metric on the top stratum of $\text{Sym}^+(p)$, but not on all of $\text{Sym}^+(p)$. In [9], we use $d_{\mathcal{SR}}$ to generate a true metric on $\text{Sym}^+(p)$.

This paper has several interrelated goals: to provide a systematic description of the geometry and topology of the triple $(M(p), F, \text{Sym}^+(p))$ in terms of fibers and strata; to lay a firm foundation for further study of the scaling-rotation framework, such as in [8] and [9]; to characterize $d_{\mathcal{SR}}$ and MSSR curves in a way that reduces computational complexity and that, for small enough p , enables the derivation of closed-form formulas; to address nonuniqueness questions for MSSR curves between two given points $X, Y \in \text{Sym}^+(p)$; and to present some unanticipated geometric results (described in more detail below), potentially of independent interest, that were produced in the course of studying the nonuniqueness questions. We elaborate below how this paper is structured relative to these goals.

Section 2 of this paper is devoted to a thorough understanding of the fibers of F . This includes a discussion of an important finite group \tilde{S}_p^+ , the group of “even signed-permutations”, that acts in an isometric and fiber-preserving way on $M(p)$, inducing a transitive action on the set of connected components of each fiber (thereby allowing us to count the connected components). To place the fibers into the context of a “big picture” view of the triple $(M(p), F, \text{Sym}^+(p))$, in Section 3 we discuss the strata of $M(p)$ and $\text{Sym}^+(p)$, and the labeling of strata by partitions. The strata and their labelings are always at work in the background of the scaling-rotation framework, even when not mentioned explicitly, since the fiber-type varies from stratum to stratum of $\text{Sym}^+(p)$, and because the more-refined stratification of $M(p)$ is needed in order to characterize the set of all MSSR curves from one point of $\text{Sym}^+(p)$ to another. The labeling by partitions plays a critical role in [8, Sections 5 and 6] and [9], for example.

With the “home space” $(M(p), F, \text{Sym}^+(p))$ of the scaling-rotation framework now described, in Section 4 we turn our attention to scaling-rotation distance and MSSR curves. In Section 4.1 we review the basics of SSR curves, before restricting attention to MSSR curves in Section 4.2 and beyond. As mentioned earlier, MSSR curves from X to Y are images of shortest-length geodesics in $M(p)$ between the submanifolds \mathcal{E}_X and \mathcal{E}_Y . The scaling-rotation distance $d_{\mathcal{SR}}(X, Y)$ is the length of any such geodesic, and we call the pair of endpoints of such a geodesic a *minimal pair*. Effectively, computing $d_{\mathcal{SR}}(X, Y)$ amounts to finding minimal pairs. However, because of the action of \tilde{S}_p^+ , minimal pairs are never unique for any (X, Y) . For some (X, Y) , there are also minimal pairs that are not related to each other by this action. *A priori*, to compute $d_{\mathcal{SR}}(X, Y)$ and to find all minimal pairs in $\mathcal{E}_X \times \mathcal{E}_Y$ entails computing all the minimal-length geodesics from each connected component of \mathcal{E}_X to each connected component of \mathcal{E}_Y . Proposition 4.10, which concludes Section 4.2, takes advantage of the \tilde{S}_p^+ -action to reduce the complexity of computing $d_{\mathcal{SR}}(X, Y)$, of finding all MSSR curves from X to Y , and of characterizing all minimal pairs. Among the outcomes is that for any connected component of \mathcal{E}_X , each MSSR curve is represented by a geodesic whose initial point lies in that component, and the number of connected components of \mathcal{E}_Y that must be considered is only the number of double-cosets of \tilde{S}_p^+ relative to two subgroups determined by X and Y .

In Section 4.3, we address uniqueness questions for MSSR curves, the most basic of which is: under what conditions on $X, Y \in \text{Sym}^+(p)$ is there a unique MSSR curve from X to Y ? For there to be more than one MSSR curve from X to Y , there must exist distinct shortest-length geodesics $\gamma_1, \gamma_2 : [0, 1] \rightarrow M(p)$ from $\mathcal{E}_X \rightarrow \mathcal{E}_Y$ such that $F \circ \gamma_1 \neq F \circ \gamma_2$. There are essentially two ways, not mutually exclusive, that this can happen: (i) there can exist such γ_i ($i = 1, 2$) whose endpoint-pairs are distinct minimal pairs, and (ii) there exist such γ_i whose endpoint-pairs are the same minimal pair. We call these possibilities “Type I” and “Type II” non-uniqueness, respectively. A minimal pair $((U, D), (V, \Lambda)) \in M(p) \times M(p)$ has more than one minimal geodesic connecting its first and second points if and only if the pair $(U, V) \in SO(p) \times SO(p)$ is *geodesically antipodal*, meaning that each of U, V is in the cut-locus of the other; equivalently, that $V^{-1}U$ is an involution. Thus, Type II nonuniqueness occurs for *some* $X, Y \in \text{Sym}^+(p)$ if and only there exists a minimal pair $((U, D), (V, \Lambda))$ for which (U, V) is geodesically antipodal. Our chief tool for determining whether such minimal pairs exist is a property we call *sign-change reducibility*: we say that the pair (U, V) is sign-change reducible if $d_{SO}(U, V)$ can be reduced by multiplying U or V by a (positive-determinant) “sign-change matrix”, a diagonal matrix each of whose diagonal entries is ± 1 . We show in Proposition 4.16 that if $(U, V) \in SO(p) \times SO(p)$ is *not* sign-change reducible, then there exist D, Λ in the top stratum of $\text{Diag}^+(p)$ such that $((U, D), (V, \Lambda))$ is a minimal pair. We show in Proposition 4.14 that for $p \leq 4$, every geodesically antipodal pair (U, V) is sign-change reducible, and that for $p \geq 11$, there exist geodesically antipodal pairs that are not sign-change reducible. From these propositions we deduce that Type II nonuniqueness never occurs for $p \leq 4$ (Corollary 4.15), and that

it always occurs for *some* $(X, Y) \in \text{Sym}^+(p) \times \text{Sym}^+(p)$ if $p \geq 11$ (Corollary 4.17). We do not believe that either of the numbers 4 and 11 above is sharp; our methods are simply not conclusive for $5 \leq p \leq 10$.

Together, Proposition 4.16 and Corollary 4.17 show that sign-change reducibility is the only obstruction to having points X, Y in the top stratum of $\text{Sym}^+(p)$ for which the set $\mathcal{M}(X, Y)$ of MSSR curves from X to Y exhibits Type II nonuniqueness.

We mention that even without Proposition 4.14, for $p \leq 3$ it is rather trivial that all geodesically antipodal pairs are sign-change reducible, and for $p = 4$ an independent proof relying on quaternions is also possible. However, our proof of the $p \leq 4$ part of Proposition 4.14 makes no use of quaternions, and unifies these low- p results.

A more refined version of the nonuniqueness question is: for each pair $(X, Y) \in \text{Sym}^+(p) \times \text{Sym}^+(p)$, what is the set $\mathcal{M}(X, Y)$? By characterizing all minimal pairs, Proposition 4.10 provides a starting point for answering this question. But to completely understand $\mathcal{M}(X, Y)$ —or even just determine its cardinality—we still need a way to tell whether MSSR curves corresponding to two (not necessarily distinct) minimal pairs with first point in a given connected component of \mathcal{E}_X are the same. In Section 4.4, Proposition 4.19 provides such a tool. While Propositions 4.10 and 4.19 are general- p results, they were originally proven in order to enable concrete computations in the case $p = 3$, the home of diffusion-tensor imaging. These are propositions from which we derive no additional conclusions in this paper, but that are crucial to [8, Sections 5 and 6]. There, aided by the quaternionic parametrization of $SO(3)$, these two results are the key tools that, for $p = 3$, allow us to find closed-form formulas for $d_{\mathcal{SR}}(X, Y)$, and to explicitly describe the set $\mathcal{M}(X, Y)$ (in particular, knowing when there is a unique MSSR curve from X to Y) when X and Y do not both lie in the top stratum.

Our proof of Proposition 4.14, completed in Section 8 after laying groundwork in Sections 5–7, takes us in some unexpected directions, with unanticipated consequences. We initially introduced the notion of sign-change reducibility into our scaling-rotation-curve study as an *ad hoc* tool to help us determine whether Type II nonuniqueness of MSSR curves is possible. However, as we show in Proposition 5.11, a refined version of the question of whether sign-change reducibility occurs is equivalent to a question purely about the geometry of Grassmannians equipped with a standard Riemannian metric: for m even and positive, is every m -dimensional subspace of \mathbf{R}^p within a certain distance $c(m)$ of a coordinate m -plane? (This question can, of course, be asked without restricting the parity of m , but the above equivalence leads us to consider only even m in this paper.) By constructing examples, we show that for $m = 2$, the answer to the latter question is no for $p \geq 11$. This, combined with the equivalence result in Proposition 5.11, yields the “ $p \geq 11$ ” part of Proposition 4.14 mentioned above. The “ $p \leq 4$ ” part of Proposition 4.14 is proven by other means (via the more technical Proposition 5.6).

Although it was the possibility of Type-II non-uniqueness for MSSR curves that led to us to the question above concerning the geometry of Grassmannians,

this question and our study of it may be of independent interest, outside of any connection to our scaling-rotation framework. This study led us to investigate several related questions concerning distances between (even-dimensional) subspaces of \mathbf{R}^p and (even-dimensional) coordinate planes not necessarily of the same dimension. Perhaps the most unexpected of these is a half-angle relation stated in Proposition 5.10 and proven in Section 6: for any two involutions $R_1, R_2 \in SO(p)$, each of the principal angles between the (-1) -eigenspaces of R_1 and R_2 is exactly half a correspondingly indexed normal-form angle of $R_1 R_2$. This relationship holds whether or not the dimensions of the (-1) -eigenspaces are equal. When the dimensions *are* equal, we use this relationship to show that a natural correspondence between $\text{Gr}_m(\mathbf{R}^p)$ and a connected component $\text{Inv}_m(p) \subset \text{Inv}(p)$, discussed in Remark 5.2, is a metric-space isometry up to a constant factor of 2 (Proposition 5.9). This isometric relation is also deducible (and may already be known) from a purely Riemannian approach, but our proof uses essentially no Riemannian geometry (see Remark 6.5 for a more precise statement, and an additional interpretation of what our proof of Proposition 5.9 shows).

The most important results coming from our study of sign-change reducibility are stated in Section 5, with the proofs deferred to Sections 6, 7, and 8. These results include those mentioned above and one more whose statement involves terminology not included in this Introduction: Proposition 5.8, a special case of a more general conjecture we make about sign-change reducibility (Conjecture 5.7). Key to almost all of these results are several facts, established in the long and technical Lemma 6.2, concerning the product of a general involution in $SO(p)$ and a positive-determinant sign-change matrix.

In this paper, there are numerous instances of a group G acting from the left on a space X . We use the notation $(g, x) \mapsto g \cdot x$ for all such actions (where $g \in G, x \in X$), and X/G will always denote the corresponding quotient space (the space of orbits, with the quotient topology). When X is a finite set, we always give X (and therefore X/G) the discrete topology.

2. Partitions and Fibers

2.1. Partitions and eigenstructure

Let $\text{Part}(\{1, \dots, p\})$ denote the set of partitions of $\{1, 2, \dots, p\}$, and $\text{Part}(p)$ the set of partitions of the integer p . We will write elements J of $\text{Part}(\{1, \dots, p\})$ are in the form $J = \{J_1, J_2, \dots\}$, where the J_i are the blocks of J . We will consider several stratified spaces in this paper. For each space, the strata are labeled naturally either by $\text{Part}(\{1, \dots, p\})$ or by $\text{Part}(p)$.

The natural left-action of the symmetric group S_p on $\{1, 2, \dots, p\}$ induces left-actions of S_p on $\text{Part}(\{1, \dots, p\})$ and \mathbf{R}^p . There is a canonical bijection between the quotient $\text{Part}(\{1, \dots, p\})/S_p$ and the set $\text{Part}(p)$, so we will implicitly regard these as the same set. For $J \in \text{Part}(\{1, \dots, p\})$, we write $[J]$ for the image of J in $\text{Part}(p)$ under the quotient map.

The sets $\text{Part}(\{1, \dots, p\})$ and $\text{Part}(p)$ are partially ordered by the refinement relation. For $J, K \in \text{Part}(\{1, \dots, p\})$, we write $J \leq K$ if K refines J . Similarly, for

$[J], [K] \in \text{Part}(p)$ we write $[J] \leq [K]$ if $[K]$ refines $[J]$. (We use this ordering, rather than its opposite, to make the bijection between partitions and strata discussed in Section 3 order-preserving.) In each of these partially ordered sets there is a well-defined “highest” (most refined) and “lowest” (least refined) element; we denote these with the subscripts “top” and “bot” respectively. Note that the quotient map $\text{Part}(\{1, \dots, p\}) \rightarrow \text{Part}(p)$ is order-preserving.

Notation 2.1. 1. Let $\text{Diag}(p)$ denote the set of $p \times p$ diagonal matrices, and let $\text{Diag}^+(p) := \{\text{diag}(d_1, \dots, d_p) : d_i > 0, 1 \leq i \leq p\} \subset \text{Diag}(p)$. For $D = \text{diag}(d_1, \dots, d_p) \in \text{Diag}(p)$, let J_D denote the partition of $\{1, 2, \dots, p\}$ determined by the equivalence relation $i \sim_D j \iff d_i = d_j$.

2. For $\emptyset \neq J \subset \{1, 2, \dots, p\}$, let $\mathbf{R}^J \subset \mathbf{R}^p$ denote the subspace $\{(x_1, \dots, x_p) \in \mathbf{R}^p \mid x_j = 0 \ \forall j \notin J\}$. For a partition $J = \{J_1, \dots, J_r\}$ of $\{1, 2, \dots, p\}$, let $\{W_1, \dots, W_r\} = \{W_1^J, \dots, W_r^J\} = \{\mathbf{R}^{J_1}, \dots, \mathbf{R}^{J_r}\}$ denote the corresponding subspaces of \mathbf{R}^p ; note that we have an orthogonal decomposition $\mathbf{R}^p = \mathbf{R}^{J_1} \oplus \dots \oplus \mathbf{R}^{J_r}$. Define the subgroup $G_J \subset SO(p)$ by

$$G_J = \{R \in SO(p) \mid RW_i = W_i, 1 \leq i \leq r\}, \quad (2.1)$$

a Lie group with (generally) more than one connected component. We write G_J^0 for the identity component of G_J .

For any subgroup $H \subset O(p)$, we write $S(H)$ for $H \cap SO(p)$. Note that

$$G_J \cong S(O(W_1) \times O(W_2) \times \dots \times O(W_r)) \quad (2.2)$$

$$\cong S(O(|J_1|) \times O(|J_2|) \cdots \times O(|J_r|)), \quad (2.3)$$

where $O(W_i)$ denotes the orthogonal group of the subspace W_i , which we identify with a subgroup of $O(p)$. Hence, writing $k_i = |J_i|$, we have

$$G_J^0 \cong SO(k_1) \times SO(k_2) \cdots \times SO(k_r). \quad (2.4)$$

Obviously (2.2) holds whether or not $k_1 \geq k_2 \geq \dots \geq k_r$. However, if the k_i are non-decreasing then $[J] = (k_1, \dots, k_r)$, so for the sake of concreteness we define

$$G_{[J]}^0 = SO(k_1) \times SO(k_2) \cdots \times SO(k_r) \quad \text{if } [J] = (k_1, k_2, \dots, k_r).$$

Remark 2.2. An easily-verified fact, reflected in the stratifications of $\text{Sym}^+(p)$ and $M(p)$ discussed in Section 3, is the following:

$$J \leq K \iff G_J \supset G_K. \quad (2.5)$$

The groups G_J are related to eigenstructure as follows. For each $D \in \text{Diag}(p)$ let G_D denote the stabilizer of D under the action of $SO(p)$ on $\text{Sym}(p)$ by conjugation, and observe that if $(U, D) \in M(p)$ is an eigen-decomposition of $X \in \text{Sym}^+(p)$, and R lies in G_D , then (UR, D) is also an eigen-decomposition of X . But G_D is precisely the group G_{J_D} defined using Notation 2.1. (In other

words, the group G_D does not depend on the absolute or relative sizes of the diagonal entries of D , but only on which eigenvalues are equal to which other eigenvalues.)

2.2. Signed permutations and signed-permutation matrices

Let $\mathcal{I}_p = (\mathbf{Z}_2)^p$ (the direct product of p copies of \mathbf{Z}_2). The role of \mathbf{Z}_2 will be as the *group of signs*, so we write it multiplicatively, with elements ± 1 . We will write typical elements of \mathcal{I}_p as $\boldsymbol{\sigma} = (\sigma_1, \dots, \sigma_p)$; we write the identity element as $\mathbf{1}$.

Both \mathcal{I}_p and the symmetric group S_p have natural representations on \mathbf{R}^p : the group \mathcal{I}_p acts via sign-changes of the coordinates, while S_p permutes the coordinates. These representations embed \mathcal{I}_p and S_p as subgroups of $O(p)$, together generating a group of “signed-permutation matrices”. Abstractly, this group is a semidirect product $\tilde{S}_p = \mathcal{I}_p \rtimes S_p$, a split extension of S_p by \mathcal{I}_p . Let $\mathbf{mat} : \tilde{S}_p \rightarrow O(p)$ denote the induced embedding of the abstract group \tilde{S}_p into $O(p)$. Writing elements of \tilde{S}_p (uniquely) as pairs $(\boldsymbol{\sigma}, \pi)$ with $\boldsymbol{\sigma} \in \mathcal{I}_p$ and $\pi \in S_p$, define $\widehat{\text{sgn}}(\boldsymbol{\sigma}, \pi) = \text{sgn}(\boldsymbol{\sigma})\text{sgn}(\pi)$, where $\text{sgn}(\sigma_1, \dots, \sigma_p) = \prod_{i=1}^p \sigma_i$ and $\text{sgn}(\pi)$ is the sign of the permutation π . The homomorphism $\widehat{\text{sgn}} : \tilde{S}_p \rightarrow \mathbf{Z}_2$ satisfies $\widehat{\text{sgn}}(\boldsymbol{\sigma}, \pi) = \det(\mathbf{mat}(\boldsymbol{\sigma}, \pi))$. Hence

$$\tilde{S}_p^+ := \ker(\widehat{\text{sgn}}), \quad (2.6)$$

which we call the group of *even* signed permutations, is exactly the set of signed permutations mapped by \mathbf{mat} into $SO(p)$; we call $\mathbf{mat}(\tilde{S}_p^+)$ the group of *even* signed-permutation matrices. We call elements of $\mathbf{mat}(\mathcal{I}_p)$ *sign-change matrices*; these are simply diagonal matrices with ± 1 's along the diagonal.

Writing $\mathcal{I}_p^+ := \{\boldsymbol{\sigma} \in \mathcal{I}_p : \text{sgn}(\boldsymbol{\sigma}) = 1\}$, we have a short exact sequence

$$1 \rightarrow \mathcal{I}_p^+ \xrightarrow{\text{incl}} \tilde{S}_p^+ \xrightarrow{\text{proj}_2} S_p \rightarrow 1. \quad (2.7)$$

Since $\mathcal{I}_p^+ \cong (\mathbf{Z}_2)^{p-1}$ (non-canonically), \tilde{S}_p^+ is an extension of S_p by $(\mathbf{Z}_2)^{p-1}$. This extension is non-split, in contrast to the extension \tilde{S}_p of S_p by $(\mathbf{Z}_2)^p$. However, from (2.7) it is immediate that

$$|\tilde{S}_p^+| = 2^{p-1}p!. \quad (2.8)$$

Henceforth we will usually not distinguish explicitly between the abstract group \tilde{S}_p (respectively \tilde{S}_p^+) the corresponding subgroup of $O(p)$ (resp. $SO(p)$); we will generally use the terms “signed permutation” and “signed-permutation matrix” synonymously in any context in which only one interpretation makes sense. However, to avoid some odd-looking formulas, for $g \in \tilde{S}_p$ (respectively $\pi \in S_p$) we will generally write P_g (resp. P_π) for the matrix $\mathbf{mat}(g)$ (resp. $\mathbf{mat}(\mathbf{1}, \pi)$), and write I_σ for $\mathbf{mat}(\boldsymbol{\sigma}, \text{id.})$. The image of g under the projection $\text{proj}_2 : \tilde{S}_p \rightarrow S_p$ will be denoted π_g .

Remark 2.3. For a subspace $W \subset \mathbf{R}^p$ and $\epsilon \in \mathbf{Z}_2 = \{\pm 1\}$, let $O_\epsilon(W) \subset O(W)$ denote the set of orthogonal transformations with determinant ϵ . In the setting of (2.2), the connected components of G_J are $O_{\epsilon_1}(W_1) \times O_{\epsilon_2}(W_2) \times \cdots \times O_{\epsilon_r}(W_r)$, subject to the restriction $\prod_i \epsilon_i = 1$. Thus a labeling of the blocks of an r -block partition J yields a 1-1 correspondence between \mathcal{I}_r^+ and the set of connected components of G_J . In particular, the number of connected components is 2^{r-1} .

Identifying $\text{Diag}(p)$ with \mathbf{R}^p , the natural left-action of S_p on \mathbf{R}^p yields a left-action of S_p on $\text{Diag}(p)$. For $D \in \text{Diag}(p)$, we will write $[D]$ for its image in the quotient space $\text{Diag}(p)/S_p$.

Note that the action of S_p on $\text{Diag}^+(p) \subset \text{Diag}(p)$ lifts to an action of \tilde{S}_p on $\text{Diag}^+(p)$,

$$g \cdot D := \pi_g \cdot D. \quad (2.9)$$

Of course, the action of \tilde{S}_p on \mathbf{R}^p induces an action on $\text{Sym}(p)$, $(g, X) \mapsto P_g X P_g^T = P_g X P_g^{-1}$. This conjugation action preserves the space $\text{Diag}^+(p)$, on which it reduces to (2.9): $P_g D P_g^{-1} = \pi_g \cdot D$.

We end this subsection by stating an easy lemma (proof left to reader) that will be used in the proof of Proposition 2.5 in the next (sub-)subsection:

Lemma 2.4. $\mathcal{I}_p^+ \subset G_J$ for all $J \in \text{Part}(\{1, \dots, p\})$. ■

2.3. Structure of the fibers

In this subsection we will provide a systematic description of the fibers of F . The starting point is the following proposition.

Proposition 2.5. Let $X \in \text{Sym}^+(p)$, $(U, D) \in \mathcal{E}_X = F^{-1}(X)$, and write $G_D = G_{J_D}$. Then

$$\mathcal{E}_X = \{(UR(P_g)^{-1}, \pi_g \cdot D) : R \in G_D, g \in \tilde{S}_p^+\}. \quad (2.10)$$

Proof: This is a simple corollary of [11, Theorem 3.3]. Details are left to the reader. ■

Corollary 2.6. Let $X \in \text{Sym}^+(p)$ and $(U, D) \in \mathcal{E}_X$, and write $G_D^0 = G_{J_D}^0$. Then

$$\mathcal{E}_X = \{(UR(P_g)^{-1}, \pi_g \cdot D) : R \in G_D^0, g \in \tilde{S}_p^+\}. \quad (2.11)$$

Proof: Clearly the right-hand side of (2.11) is contained in the right-hand side of (2.10), so it suffices to prove the opposite inclusion.

Let $R \in G_D, g \in \tilde{S}_p^+$. Enumerate the blocks of $J := J_D$ as J_1, \dots, J_r , and let W_i be as in Notation 2.1. As noted in Remark 2.3, the enumeration of the blocks of J yields a 1-1 correspondence between \mathcal{I}_r^+ and the connected components of G_J . Let R lie in the component of G_J labeled by $(\epsilon_1, \dots, \epsilon_r) \in \mathcal{I}_r^+$. The

cardinality of $\{j : \epsilon_j = -1\}$ is some even number k . Let $\sigma = (\sigma_1, \dots, \sigma_p) \in \mathcal{I}_p$, where for $1 \leq i \leq p$ we set

$$\sigma_i = \begin{cases} -1 & \text{if } i \in J_j, \epsilon_j = -1, \text{ and } i \text{ is the smallest element of } J_j ; \\ 1 & \text{otherwise.} \end{cases}$$

Then $R_1 := I_\sigma R \in G_J^0$. But also $|\{i : \sigma_i = -1\}| = k$, so $\sigma \in \mathcal{I}_p^+ \subset \tilde{S}_p^+$, and $P_g I_\sigma = P_{g_1}$ for some $g_1 \in \tilde{S}_p^+$ with $\pi_{g_1} = \pi_g$. Hence $P_g R = (P_g I_\sigma)(I_\sigma R) = P_{g_1} R_1$, so $(U(P_g R)^{-1}, \pi_g \cdot D) = (U(P_{g_1} R_1)^{-1}, \pi_{g_1} \cdot D)$, which lies in the right-hand side of (2.11). The desired inclusion follows. \blacksquare

We will need a bit more notation to complete our characterization of the fibers of F :

Notation 2.7.

1. For $J = \{J_1, \dots, J_r\} \in \text{Part}(\{1, \dots, p\})$, let $K_J = \{\pi \in S_p : \pi(J_i) = J_i, 1 \leq i \leq r\}$, $\Gamma_J = \tilde{S}_p^+ \cap G_J$, and $\Gamma_J^0 = \Gamma_J \cap G_J^0 = \tilde{S}_p^+ \cap G_J^0$. Also define the subgroup

$$\mathcal{I}_J^+ = \{(\sigma_1, \dots, \sigma_p) \in \mathcal{I}_p : \prod_{j \in J_i} \sigma_j = 1, 1 \leq i \leq r\} \subset \mathcal{I}_p^+. \quad (2.12)$$

2. For any $X \in \text{Sym}^+(p)$ and $(U, D) \in \mathcal{E}_X$, define

$$[(U, D)] = \{(UR, D) : R \in G_D^0\}, \quad (2.13)$$

the connected component of \mathcal{E}_X containing (U, D) . We write $\text{Comp}(\mathcal{E}_X)$ for the set of connected components of \mathcal{E}_X .

3. For any Lie group G and closed subgroup K , we write G/K and $K \backslash G$ for the spaces of left- and right-cosets, respectively, of K in G . (In particular, we use this notation when G is a finite group.)

Note that the groups \mathcal{I}_J^+ are a generalization of \mathcal{I}_p^+ ; for $J = J_{\text{top}} = \{\{1\}, \{2\}, \dots, \{p\}\}$ we have $\mathcal{I}_J^+ = \mathcal{I}_p^+$.

We observe that equivalent definitions of Γ_J and Γ_J^0 are:

$$\Gamma_J = \{(\sigma, \pi) \in \tilde{S}_p^+ : \pi \in K_J\}, \quad (2.14)$$

$$\Gamma_J^0 = \{(\sigma, \pi) \in \tilde{S}_p^+ : \sigma \in \mathcal{I}_J^+, \pi \in K_J\}. \quad (2.15)$$

Thus, analogously to (2.7), we have short exact sequences

$$1 \rightarrow \mathcal{I}_p^+ \rightarrow \Gamma_J \rightarrow K_J \rightarrow 1, \quad (2.16)$$

$$1 \rightarrow \mathcal{I}_J^+ \rightarrow \Gamma_J^0 \rightarrow K_J \rightarrow 1, \quad (2.17)$$

and the cardinality of the middle group in each of these sequences is the product of the cardinalities of the groups to its left and right. Note also that equivalent definitions of K_J are

$$\begin{aligned} K_J &= \{\pi \in S_p \mid \pi \cdot D = D \text{ for some } D \text{ with } J_D = J\} \\ &= \{\pi \in S_p \mid \pi \cdot D = D \text{ for all } D \text{ with } J_D = J\}. \end{aligned} \quad (2.18)$$

Next, observe that the group \tilde{S}_p^+ acts on $M(p)$ via setting

$$g \cdot (U, D) = (UP_g^{-1}, g \cdot D). \quad (2.19)$$

This action preserves every fiber of F . Thus for each $X \in \text{Sym}^+(p)$ there is an induced action of \tilde{S}_p^+ on $\text{Comp}(\mathcal{E}_X)$, given by

$$g \cdot [(U, D)] = [g \cdot (U, D)]. \quad (2.20)$$

This leads us to:

Proposition 2.8. *Let $X \in \text{Sym}^+(p)$. Then every $(U, D) \in \mathcal{E}_X$ determines a bijection between $\text{Comp}(\mathcal{E}_X)$ and the set $\tilde{S}_p^+/\Gamma_{J_D}^0$.*

Proof: Two elements $(U, D), (U', D')$ lie in the same component of \mathcal{E}_X if and only if and only if $D' = D$ and $U' = UR$ for some $R \in G_D^0$. Thus it is clear from (2.11) that the action (2.20) of \tilde{S}_p^+ on $\text{Comp}(\mathcal{E}_X)$ is transitive. Therefore for any $(U, D) \in \mathcal{E}_X$, the map $\tilde{S}_p^+ \rightarrow \text{Comp}(\mathcal{E}_X), g \mapsto g \cdot [(U, D)]$, induces a bijection $\tilde{S}_p^+/\text{Stab}([(U, D)]) \rightarrow \text{Comp}(\mathcal{E}_X)$, where $\text{Stab}([(U, D)])$ is the stabilizer of $[(U, D)]$ under the action (2.20). But, as is easily checked, $\text{Stab}([(U, D)])$ is exactly the group $\Gamma_{J_D}^0$. ■

An important special case of Proposition 2.8 is the case in which all eigenvalues of X are distinct. In this case, $J_D = J_{\text{top}} = \{\{1\}, \{2\}, \dots, \{p\}\}$ and $\Gamma_{J_D}^0 = \{\text{id.}\}$. Thus the action of \tilde{S}_p^+ on $\text{Comp}(\mathcal{E}_X)$ is free as well as transitive. Furthermore $G_D^0 = \{I\}$, so each connected component of \mathcal{E}_X is a single point; $\text{Comp}(\mathcal{E}_X) = \mathcal{E}_X$. Thus \mathcal{E}_X itself is an orbit of \tilde{S}_p^+ , and any choice of $(U, D) \in \mathcal{E}_X$ yields a bijection $\tilde{S}_p^+ \rightarrow \mathcal{E}_X, g \mapsto g \cdot (U, D)$.

Corollary 2.9. *Let $X \in \text{Sym}^+(p)$, $(U, D) \in \mathcal{E}_X$, and let k_1, \dots, k_r be the parts of the partition $[J_D]$ of p . Then \mathcal{E}_X is diffeomorphic to a disjoint union of $2^{r-1} \frac{p!}{k_1! k_2! \dots k_r!}$ copies of $SO(k_1) \times SO(k_2) \times \dots \times SO(k_r)$.*

Proof: Let $J = J_D$. It is clear from (2.13) that each connected component of \mathcal{E}_X is a submanifold of $M(p)$ diffeomorphic to $G_D^0 = G_J^0$, which from (2.4) is isomorphic (hence diffeomorphic) to $SO(k_1) \times SO(k_2) \times \dots \times SO(k_r)$. From Proposition 2.8, the number of connected components is $|\tilde{S}_p^+/\Gamma_{J_D}^0| = |\tilde{S}_p^+|/|\Gamma_{J_D}^0|$.

From (2.8) we have $|\tilde{S}_p^+| = 2^{p-1}p!$, while from (2.17) we have $|\Gamma_{J_D}^0| = |\mathcal{I}_J^+||K_J|$. It is easily seen that \mathcal{I}_J^+ is isomorphic to $(\mathbf{Z}_2)^{p-r}$, and that K_J is isomorphic to $S_{k_1} \times S_{k_2} \times \cdots \times S_{k_r}$, and hence that $|K_J| = k_1!k_2! \cdots k_r!$. The result follows. ■

Remark 2.10. An alternate, instructive route to Corollary 2.9 is the following. (We merely sketch the ideas; the reader may fill in the details.) For $J \in \text{Part}(\{1, \dots, p\})$, define $\mathcal{Q}_J = \{P_g R : g \in \tilde{S}_p^+, R \in G_J\} \subset SO(p)$. Thus the set \mathcal{Q}_J is a finite union of left-cosets of G_J , each of which is diffeomorphic to the compact submanifold $G_J \subset SO(p)$. If $X \in \text{Sym}^+(p)$, $(U, D) \in \mathcal{E}_X$, and $J = J_D$, the map $\mathcal{Q}_J \rightarrow M(p)$, $Q \mapsto (UQ^{-1}, QDQ^{-1})$, is an embedding with image \mathcal{E}_X . Hence \mathcal{E}_X is a submanifold of $M(p)$ diffeomorphic to \mathcal{Q}_J . But for any closed subgroups H_1, H_2 of a compact Lie group G , the set $H_1 H_2 := \{h_1 h_2 : h_1 \in H_1, h_2 \in H_2\} \subset G$ is a submanifold of G and a principal H_2 -bundle over $H_1/(H_1 \cap H_2)$, with projection map given by $h_1 h_2 \mapsto h_1(H_1 \cap H_2)$. Applying this to the case $H_1 = \tilde{S}_p^+, H_2 = G_J, G = SO(p)$, we have $H_1 \cap H_2 = \Gamma_J$, so \mathcal{Q}_J is a principal G_J -bundle over the *finite* set \tilde{S}_p^+/Γ_J . But the natural map $\tilde{S}_p^+/\Gamma_J \rightarrow S_p/K_J$, $g\Gamma_J \mapsto \text{proj}_2(g)K_J$ (where proj_2 is as in (2.7)), is a bijection, so \mathcal{Q}_J may be viewed as a principal G_J -bundle over S_p/K_J . The cardinality of this base-space is $|S_p|/|K_J|$, which is simply the multinomial coefficient $\frac{p!}{k_1!k_2! \cdots k_r!}$ if $[J] = (k_1, \dots, k_r) \in \text{Part}(p)$. Thus \mathcal{E}_X is diffeomorphic to $\frac{p!}{k_1!k_2! \cdots k_r!}$ copies of G_J , and each copy of G_J is diffeomorphic to 2^{r-1} copies of $SO(k_1) \times \cdots \times SO(k_r)$.

3. Stratification of $\text{Sym}^+(p)$, $M(p)$, and related spaces

The group $G = SO(p)$ acts from the left on $\text{Sym}^+(p)$ via

$$(U, X) \mapsto U \cdot X = UXU^T. \quad (3.1)$$

As for any group-action, elements $X, Y \in \text{Sym}^+(p)$ are said to have the same *orbit type* if their stabilizers are conjugate; this implies that the fibers $\mathcal{E}_X, \mathcal{E}_Y$ are diffeomorphic. The orbit-type stratification of any manifold under the action of compact Lie group is known to be a Whitney stratification ([6, p. 21]). The orbit-type stratification of $\text{Sym}^+(p)$ is easily seen to be identical to the stratification by “eigenvalue-multiplicity type” defined below.

We use $\text{Part}(\{1, \dots, p\})$ to define stratifications of the spaces $\text{Diag}^+(p)$ and $M(p)$, and use $\text{Part}(p)$ to define stratifications of $\text{Diag}^+(p)/S_p$ and $\text{Sym}^+(p)$. The commutative diagram in Figure 1 indicates the relationships among these spaces and label-sets. In this diagram, the maps not yet defined are as follows:

- $\text{proj}_2 : M(p) = SO(p) \times \text{Diag}^+(p) \rightarrow \text{Diag}^+(p)$ is projection onto the second factor.
- For $X \in \text{Sym}^+(p)$, if $(U, D) \in \mathcal{E}_X$ we define $\overline{\text{proj}_2}(X) = [D] \in \text{Sym}^+(p)/S_p$; this is well-defined since in any two eigen-decompositions X , the diagonal parts differ at most by a permutation.

$$\begin{array}{ccccc}
M(p) & \xrightarrow{\text{proj}_2} & \text{Diag}^+(p) & \xrightarrow{\text{lbl}} & \text{Part}(\{1, \dots, p\}) \\
\downarrow F & & \downarrow \text{quo}_1 & & \downarrow \text{quo}_2 \\
\text{Sym}^+(p) & \xrightarrow[\text{proj}_2]{} & \text{Diag}^+(p)/S_p & \xrightarrow[\overline{\text{lbl}}]{} & \text{Part}(p)
\end{array}$$

Figure 1: Commutative diagram defining the stratifications of $\text{Sym}^+(p)$ and related spaces.

- The stratum-labeling maps $\text{lbl}, \overline{\text{lbl}}$ are defined by $\text{lbl}(D) = J_D$, $\text{lbl}([D]) = [J_D]$.
- quo_1 and quo_2 are the quotient maps $\text{Diag}^+(p) \rightarrow \text{Diag}^+(p)/S_p$ and $\text{Part}(\{1, \dots, p\}) \rightarrow \text{Part}(\{1, \dots, p\})/S_p = \text{Part}(p)$ respectively.

We define strata as the diagram suggests: for $J \in \text{Part}(\{1, \dots, p\})$ and $[K] \in \text{Part}(p)$, (i) $\mathcal{D}_J := \text{lbl}^{-1}(J) \subset \text{Diag}^+(p)$, (ii) $\mathcal{D}_{[K]} := \overline{\text{lbl}}^{-1}([K]) \subset \text{Diag}^+(p)/S_p$, (iii) $\mathcal{S}_J := \text{proj}_2^{-1}(\mathcal{D}_J) = SO(p) \times \mathcal{D}_J \subset M(p)$, and (iv) $\mathcal{S}_{[K]} := \overline{\text{proj}_2}^{-1}(\mathcal{D}_{[K]})$. For $X \in \mathcal{S}_{[K]}$, we call the partition $[K] \in \text{Part}(p)$ the *eigenvalue-multiplicity type* of X .

It is clear that each of the four spaces $\text{Diag}^+(p)$, $M(p)$, $\text{Diag}^+(p)/S_p$, and $\text{Sym}^+(p)$, is the disjoint union of the strata we have defined in that space. Note also that for any $J \in \text{Part}(\{1, \dots, p\})$, $F(\mathcal{S}_J) = \mathcal{S}_{[J]}$ and $\text{quo}_1(\mathcal{D}_J) = \mathcal{D}_{[J]}$. One can easily check that

$$\overline{\mathcal{D}_K} = \bigcup_{J \leq K} \mathcal{D}_J, \quad \overline{\mathcal{S}_K} = \bigcup_{J \leq K} \mathcal{S}_J, \quad (3.2)$$

$$\overline{\mathcal{D}_{[K]}} = \bigcup_{[J] \leq [K]} \mathcal{D}_{[J]}, \quad \overline{\mathcal{S}_{[K]}} = \bigcup_{[J] \leq [K]} \mathcal{S}_{[J]}. \quad (3.3)$$

In any stratified space, there is a natural partial ordering “ \leq ” on the set of strata \mathcal{T}_i defined by declaring $\mathcal{T}_1 \leq \mathcal{T}_2$ if $\mathcal{T}_1 \subset \overline{\mathcal{T}_2}$. Then (3.2)–(3.3) imply that for all $J, K \in \text{Part}(\{1, \dots, p\})$, in the appropriate spaces we have

$$\mathcal{S}_J \leq \mathcal{S}_K \iff J \leq K \iff \mathcal{D}_J \leq \mathcal{D}_K, \quad (3.4)$$

$$\mathcal{S}_{[J]} \leq \mathcal{S}_{[K]} \iff [J] \leq [K] \iff \mathcal{D}_{[J]} \leq \mathcal{D}_{[K]}. \quad (3.5)$$

In each of the stratified spaces above, there is a highest stratum and a lowest stratum, namely the strata labeled by $J_{\text{top}}, [J_{\text{top}}], J_{\text{bot}}$, and $[J_{\text{bot}}]$. Note that for $J, K \in \text{Part}(\{1, \dots, p\})$, $J \leq K$ implies $[J] \leq [K]$. In view of (3.4)–(3.5), a similar comment applies to our stratified spaces: $\mathcal{S}_J \leq \mathcal{S}_K$ implies $\mathcal{S}_{[J]} \leq \mathcal{S}_{[K]}$, and $\mathcal{D}_J \leq \mathcal{D}_K$ implies $\mathcal{D}_{[J]} \leq \mathcal{D}_{[K]}$. The converse of each of these implications is false for $p > 2$.

4. Scaling-rotation distance and minimal smooth scaling-rotation curves

The Lie groups $SO(p)$ and $\text{Diag}^+(p)$ carry natural bi-invariant Riemannian metrics. If we endow $M(p) = SO(p) \times \text{Diag}^+(p)$ with a product Riemannian metric g_M , the geodesics γ in $(M(p), g_M)$ are easily computed. We define *smooth scaling-rotation (SSR) curves* in $\text{Sym}^+(p)$ to be the projections to $\text{Sym}^+(p)$ of the geodesics in $(M(p), g_M)$, i.e. curves of the form $F \circ \gamma$. (In [13] and [11] these were called simply “scaling-rotation curves”. Remark 4.8 explains why we have added “smooth” to this name.)

4.1. Smooth scaling-rotation curves

The Lie algebra $\mathfrak{so}(p) = T_I(SO(p))$ is the space of $p \times p$ antisymmetric matrices. For $U \in SO(p)$, the tangent space $T_U(SO(p))$ can be identified with either the left-translate or right-translate of $\mathfrak{so}(p)$ by U . For the purposes of this paper the latter identification is somewhat more convenient:

$$T_U(SO(p)) = \{AU : A \in \mathfrak{so}(p)\}. \quad (4.1)$$

With this identification, the standard bi-invariant Riemannian metric g_{SO} on $SO(p)$ is defined by

$$g_{SO}|_U(A_1, A_2) = -\frac{1}{2}\text{tr}(A_1 U^{-1} A_2 U^{-1}), \quad (4.2)$$

where $U \in SO(p)$ and $A_1, A_2 \in T_U(SO(p))$. The requirement of bi-invariance determines the metric g_{SO} up to a constant factor unless $p = 4$. This is true for trivial reasons if $p = 2$; for larger $p \neq 4$ it follows from the simplicity of $\mathfrak{so}(p)$. For $p = 4$ the isomorphism $\mathfrak{so}(4) \cong \mathfrak{so}(3) \oplus \mathfrak{so}(3)$ implies that there is a two-parameter family of bi-invariant metrics; for this p the choice (4.2) amounts to fixing two scaling-parameters rather than one.)

The manifold $\text{Diag}^+(p)$ is also a Lie group, but since it is an open subset of the vector space $\text{Diag}(p)$ we make the identification $T_D(\text{Diag}^+(p)) = \text{Diag}(p)$ for all $D \in \text{Diag}^+(p)$. This abelian Lie group carries a (bi-)invariant Riemannian metric $g_{\mathcal{D}^+}$ defined by

$$g_{\mathcal{D}^+}|_D(L_1, L_2) = \text{tr}(D^{-1} L_1 D^{-1} L_2) \quad (4.3)$$

where $D \in \text{Diag}^+(p)$ and $L_1, L_2 \in T_D(\text{Diag}^+(p))$. There is a p -parameter family of bi-invariant metrics on $\text{Diag}^+(p)$, but, up to a constant factor, $g_{\mathcal{D}^+}$ is the unique one that is also invariant under the action of the symmetric group S_p .

On $M(p)$, we will use a product Riemannian metric determined by the metrics on the two factors $SO(p), \text{Diag}^+(p)$. Specifically, letting $k > 0$ be an arbitrary parameter that can be chosen as desired for applications, we set

$$g_M|_{(U,D)}((A_1, L_1), (A_2, L_2)) = k g_{SO}|_U(A_1, A_2) + g_{\mathcal{D}^+}|_D(L_1, L_2), \quad (4.4)$$

where $U \in SO(p), D \in \text{Diag}^+(p)$, and $(A_i, L_i) \in T_U(SO(p)) \oplus T_D(\text{Diag}^+(p)) = T_{(U,D)}M(p)$. Since the metrics g_{SO} and $g_{\mathcal{D}^+}$ are bi-invariant, the geodesics in

$M(p)$ can be obtained as either left-translates or right-translates of geodesics through the identity (I, I) . In this paper, the latter will be more convenient.

Definition 4.1. A *smooth scaling-rotation (SSR) curve* is a curve χ in $\text{Sym}^+(p)$ of the form $F \circ \gamma$, where $\gamma : I \rightarrow M(p)$ is a geodesic defined on some interval I .

Note: in this paper, we use *curve* sometimes to mean a *parametrized curve*—a map $I \rightarrow Z$, where I is an interval and Z is some space—and sometimes to mean an equivalence class of such maps, where two maps are regarded as equivalent if one is a monotone reparametrization of the other. (In particular, two such equivalent maps have the same image.) It should always be clear from context which meaning is intended.

Notation 4.2. For $(U, D) \in M(p)$, $A \in \mathfrak{so}(p)$, and $L \in \text{Diag}(p)$, we define $\gamma_{U,D,A,L} : \mathbf{R} \rightarrow M(p)$ and $\chi_{U,D,A,L} : \mathbf{R} \rightarrow \text{Sym}^+(p)$ by

$$\gamma_{U,D,A,L}(t) = (\exp(tA)U, \exp(tL)D) \quad (4.5)$$

and

$$\chi_{U,D,A,L} = F \circ \gamma_{U,D,A,L}. \quad (4.6)$$

We use the same notation $\gamma_{U,D,A,L}$, $\chi_{U,D,A,L}$ for the restrictions of the curves above to any interval $I \subset \mathbf{R}$.

The curve $\gamma_{U,D,A,L} : \mathbf{R} \rightarrow M(p)$ is the geodesic in $M(p)$ with initial conditions $\gamma(0) = (U, D)$, $\gamma'(0) = (AU, DL) \in T_{(U,D)}M(p)$. The projection of $\gamma_{U,D,A,L}$ to $\text{Sym}^+(p)$, the curve $\chi_{U,D,A,L}$, is the corresponding smooth scaling-rotation curve.

It well known that in the Riemannian manifold $(SO(p), g_{SO})$, the cut-locus of the identity is the set of all involutions, $\{R \in SO(p) \mid R^2 = I \neq R\}$. For every non-involution $R \in SO(p)$, there is a unique $A \in \mathfrak{so}(p)$ of smallest norm such that $\exp(A) = R$ (see Section 5.1); we define $\log(R) = A$. If R is an involution, there is more than one smallest-norm $A \in \mathfrak{so}(p)$ such that $\exp(A) = R$, and we allow $\log(R)$ to denote the *set* of all such A 's. However, all elements A in this set have the same norm, which we write as $\|\log(R)\|$. Thus $\|\log(R)\|$ is a well-defined real number for all $R \in SO(p)$, even though $\log(R)$ is not a unique element of $\mathfrak{so}(p)$ when R is an involution. With this understood, the geodesic-distance function d_M on $M(p)$ is given by

$$d_M^2((U, D), (V, \Lambda)) = k d_{SO}(U, V)^2 + d_{\mathcal{D}^+}(D, \Lambda)^2 \quad (4.7)$$

$$= \frac{k}{2} \|\log(U^{-1}V)\|^2 + \|\log(D^{-1}\Lambda)\|^2, \quad (4.8)$$

where in (4.8) and for the rest of this paper, $\|\cdot\|$ denotes the Frobenius norm on matrices: $\|A\|^2 = \|A\|_F^2 = \text{tr}(A^T A)$ for any matrix A .

The invariances of the metrics d_{SO} and $d_{\mathcal{D}^+}$ lead to the following:

Proposition 4.3 ([11, Proposition 3.7]). *The geodesic distance (4.7) on $M(p)$ is invariant under simultaneous left or right multiplication by orthogonal matrices, permutations and scaling: For any $R_1, R_2 \in O(p)$, $\pi \in S_p$ and $S \in \text{Diag}^+(p)$, and for any $(U, D), (V, \Lambda) \in (SO \times \text{Diag}^+)(p)$,*

$$d_M((U, D), (V, \Lambda)) = d_M((R_1 U R_2, S\pi \cdot D), (R_1 V R_2, S\pi \cdot \Lambda)).$$

4.2. Scaling-rotation distance and MSSR curves

Definition 4.4 ([11, Definition 3.10]). For $X, Y \in \text{Sym}^+(p)$, the *scaling-rotation distance* $d_{\mathcal{SR}}(X, Y)$ between X and Y is defined by

$$d_{\mathcal{SR}}(X, Y) := \inf_{\substack{(U, D) \in \mathcal{E}_X, \\ (V, \Lambda) \in \mathcal{E}_Y}} d_M((U, D), (V, \Lambda)). \quad (4.9)$$

Definition 4.5. Let γ be a piecewise-smooth curve in $M(p)$ and let $\ell(\gamma)$ denote the length of γ . For $X, Y \in \text{Sym}^+(p)$, we call $\gamma : [0, 1] \rightarrow M(p)$ an *F-minimal geodesic* (from \mathcal{E}_X to \mathcal{E}_Y) if $\gamma(0) \in \mathcal{E}_X, \gamma(1) \in \mathcal{E}_Y$, and $\ell(\gamma) = d_{\mathcal{SR}}(X, Y)$. We call a pair of points $((U, D), (V, \Lambda)) \in \mathcal{E}_X \times \mathcal{E}_Y$ a *minimal pair* if $(U, D) = \gamma(0)$ and $(V, \Lambda) = \gamma(1)$ for some *F-minimal geodesic* γ . A *minimal smooth scaling-rotation* (MSSR) curve from X to Y is a curve χ in $\text{Sym}^+(p)$ of the form $F \circ \gamma$ where γ is an *F-minimal geodesic*. We say that the MSSR curve $\chi = F \circ \gamma$ *corresponds to* the minimal pair formed by the endpoints of γ . We let $\mathcal{M}(X, Y)$ denote the set of MSSR curves from X to Y .

Obviously an *F-minimal geodesic* is a minimal geodesic in the usual sense of Riemannian geometry: it is a curve of shortest length among all piecewise-smooth curves with the same endpoints. (From the general theory of geodesics, the image of any such curve γ is actually *smooth*.) Thus a definition equivalent to (4.9) is

$$d_{\mathcal{SR}}(X, Y) = \inf \{ \ell(\gamma) \mid \gamma : [0, 1] \rightarrow M(p) \text{ is a geodesic with } \gamma(0) \in \mathcal{E}_X, \gamma(1) \in \mathcal{E}_Y \}. \quad (4.10)$$

Thus an *F-minimal geodesic* can alternatively be defined as a geodesic of minimal length among all geodesics starting in one given fiber and ending in another.

From Corollary 2.9, every fiber of F is compact, so the infimum in (4.9) is always achieved. Hence for all $X, Y \in \text{Sym}^+(p)$, there always exists an *F-minimal geodesic*, a minimal pair in $\mathcal{E}_X \times \mathcal{E}_Y$, and an MSSR curve from X to Y . We discuss uniqueness issues in the next subsection.

Remark 4.6. Observe that we have not defined a Riemannian metric on $\text{Sym}^+(p)$, so there is no “automatic” meaning attached to the phrase *length of a smooth curve* in $\text{Sym}^+(p)$. However, for an SSR curve χ in $\text{Sym}^+(p)$ we define the *length of χ* to be $\ell(\chi) := \inf \{ \ell(\gamma) : \gamma \text{ is a geodesic in } M(p) \text{ and } F \circ \gamma = \chi \}$. With this definition, (4.10) becomes

$$d_{\mathcal{SR}}(X, Y) = \inf \{ \ell(\chi) \mid \chi : [0, 1] \rightarrow \text{Sym}^+(p) \text{ is an SSR curve with } \chi(0) = X, \chi(1) = Y \}. \quad (4.11)$$

In the term “smooth scaling-rotation curve”, the word “smooth” needs some comment. By its definition, every SSR curve $\chi : I \rightarrow \text{Sym}^+(p)$ is a smooth *map*, but it is not clear whether the image of χ is “geometrically smooth”, i.e. locally (in I) a smooth submanifold or submanifold-with-boundary of $\text{Sym}^+(p)$. For the image of χ to be “*geometrically smooth*”, χ must admit a *regular* parametrization, one that is an immersion. It turns out that all SSR curves do, except for those whose images are single points:

Proposition 4.7. *If γ is a non-constant geodesic, then $F \circ \gamma$ is either an immersion or a constant map.*

Proof: Let $\gamma : [0, 1] \rightarrow M(p)$ be a non-constant F -minimal geodesic and let $\chi = F \circ \gamma$.

Let $(U, D) = \gamma(0)$ and let $X = \chi(0) = F(U, D)$. Since γ is a geodesic there exist unique $A \in \mathfrak{so}(p)$, $L \in \text{Diag}(p)$ such that $\gamma(t) = (e^{tA}U, e^{tL}D)$. Non-constancy implies $(A, L) \neq (0, 0)$. Direct computation yields

$$\chi'(t) = e^{tA} \{ [A, U\Lambda(t)U^T] + U\Lambda(t)U^T \} e^{-tA},$$

where $\Lambda(t) = e^{tL}D$ and $[,]$ denotes matrix commutator.

Suppose that $t_0 \in [0, 1]$ is such that $\chi'(t_0) = 0$. Then

$$[A, U\Lambda(t_0)U^T] + U\Lambda(t_0)U^T = 0. \quad (4.12)$$

Multiplying on left by U^T and on the right by U yields $[\tilde{A}, \Lambda(t_0)] + L\Lambda(t_0) = 0$, where $\tilde{A} = U^T A U$. But because $\Lambda(t_0)$ is diagonal, the diagonal entries of any commutator $[B, \Lambda(t_0)]$ are zero. Since $L\Lambda(t_0)$ is a diagonal matrix, this implies that $[\tilde{A}, \Lambda(t_0)] = 0 = L\Lambda(t_0)$. But $\Lambda(t_0)$ is invertible, so the second equality implies $L = 0$. Thus $\Lambda(t) = D$ for all t , and plugging this into (4.12) with $t = t_0$ we find $[A, X] = 0$. It follows that X commutes with e^{tA} for every t . Hence $\chi(t) = e^{tA}U D U^T e^{-tA} = e^{tA}X e^{-tA} = X$ for all t .

Thus either $\chi'(t)$ is nonzero for every $t \in [0, 1]$ or χ is constant. ■

It seems likely that a non-constant MSSR curve χ is actually an embedding (for this, it suffices that χ be injective, since $[0, 1]$ is compact), but we have not proven this. There do exist *non-minimal* non-constant SSR curves that are not one-to-one. One example is any nonconstant periodic SSR curve: in (4.6), take $L = 0$ and take A to be any nonzero element of $\mathfrak{so}(p)$ for which $\exp(t_1 A) = I$ for some $t_1 \neq 0$. (For $p \leq 3$, the latter condition is redundant.) In this case, the curve $\chi_{U,D,A,L}|_{[0,|t_1|]}$ is an SSR curve of positive length from (U, D) to (U, D) . A

nonperiodic example with $p = 2$ is the following. Let $J = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$, $U(t) = \exp(t \frac{\pi}{2} J)$, $D(t) = \begin{pmatrix} e^{1-t} & 0 \\ 0 & e^t \end{pmatrix}$. Then the curve $t \mapsto \gamma(t) := (U(t), D(t))$ is the geodesic $\gamma_{U,D,A,L}$, where $U = U(0) = I$, $D = D(0) = \text{diag}(e, 1)$, $A = \frac{\pi}{2} J$, and $L = \text{diag}(-1, 1)$. Let χ be the SSR curve $F \circ \gamma$. Then, as the reader may check, if $t_1 < t_2$ we have $\chi(t_1) = \chi(t_2)$ if (and only if) for some integer $n \geq 0$

we have $t_1 = -n$ and $t_2 = n + 1$. Now let n_1, n_2 be non-negative integers, let $t_1 \in (-n_1 - 1, -n_1)$, $t_2 \in (n_2 + 1, n_2 + 2)$, and let $n = \min\{n_1, n_2\}$, $X = \chi(t_1)$, and $Y = \chi(t_2)$. Then $\chi|_{[t_1, t_2]}$ is an SSR curve from X to Y with $n + 1$ self-crossings. Note that the presence of self-crossings does not directly imply that $\chi|_{[t_1, t_2]}$ is not an MSSR curve: if we remove the closed curve $\chi|_{[-n, n+1]}$ from $\chi|_{[t_1, t_2]}$, the piecewise-smooth curve χ_1 from X to Y that remains is not an SSR curve. (As the reader may check, the set $\{\chi'(-n), \chi'(n+1)\}$ is linearly independent, so χ_1 cannot be reparametrized as an immersion. Hence, by Proposition 4.7, there is no geodesic γ_1 in $M(2)$ such that χ_1 can be reparametrized as $F \circ \gamma_1$.) Hence χ_1 is not a candidate for an SSR curve from X to Y that is shorter than χ . However, with a little effort one can check by direct computation that there is an F -minimal geodesic from X to Y that is shorter than $\gamma|_{[t_1, t_2]}$. (One can compute the length of the minimal geodesic from any of the four points in \mathcal{E}_X to any of the four points in \mathcal{E}_Y , and see that each of these lengths is less than $\ell(\gamma|_{[t_1, t_2]})$.)

Remark 4.8. As noted in [11], the “scaling-rotation distance” $d_{\mathcal{SR}}$ is *not* a metric on $\text{Sym}^+(p)$; it does not satisfy the triangle inequality. In [9], we show that the pseudometric $\rho_{\mathcal{SR}}$ generated by the semimetric $d_{\mathcal{SR}}$ is a true metric $\rho_{\mathcal{SR}}$ on $\text{Sym}^+(p)$. (It is not trivial to show that $\rho_{\mathcal{SR}}(X, Y) \neq 0$ for $X \neq Y$.) Effectively, the construction enlarges the class of scaling-rotation (SR) curves χ considered in (4.11) from smooth maps to piecewise-smooth maps (with $\ell(\chi)$ redefined correspondingly). It is for this reason we have made “smooth” part of the terminology used in Definition 4.1. This definition of the scaling-rotation metric $\rho_{\mathcal{SR}}$ is analogous to the definition of “distance between two points in a Riemannian manifold”: the infimum of the lengths of *piecewise-smooth* curves joining the points. Some minimal-length scaling-rotation curves are *geometrically* non-smooth (having corners); an MSSR curve from X to Y has minimal length only among *smooth* SR curves from X to Y . This phenomenon does not occur in Riemannian geometry; in a Riemannian manifold, minimal piecewise-smooth curves between two points are actually *smooth*, geometrically.

Definition 4.9. Recall that given any group G and subgroups H_1, H_2 , an (H_1, H_2) *double-coset* is an equivalence class under the equivalence relation \sim on G defined by declaring $g_1 \sim g_2$ if there exist $h_1 \in H_1, h_2 \in H_2$ such that $g_2 = h_1 g_1 h_2$. The set of equivalence classes under this relation is denoted $H_1 \backslash G / H_2$. By a *set of representatives* of $H_1 \backslash G / H_2$ we mean a subset of G consisting of exactly one element from each (H_1, H_2) double-coset. Note that every ordinary left coset or right coset is also a double coset (with the second subgroup taken to be trivial), so “set of representatives” is defined for ordinary cosets as well.

In [8] we apply the next proposition to compute all scaling-rotation distances, and to help compute and classify all smooth minimal scaling-rotation curves, in the case $p = 3$.

Proposition 4.10. *Let $X, Y \in \text{Sym}^+(p)$ and let $(U, D) \in \mathcal{E}_X, (V, \Lambda) \in \mathcal{E}_Y$. Let Z be any set of representatives of $\Gamma_{JD}^0 \backslash \tilde{S}_p^+ / \Gamma_{JA}^0$. Then the scaling-rotation distance from X to Y is given by*

$$d_{SR}(X, Y)^2 = \min \{ kd_{SO}(UR_U, VR_V P_g^{-1})^2 + d_{D^+}(D, \pi_g \cdot \Lambda) \}^2 : \\ R_U \in G_D^0, R_V \in G_\Lambda^0, \text{ and } g \in Z \} \quad (4.13)$$

$$= \min_{g \in Z} \left\{ k \left(\min \{ d_{SO}(UR_U, VR_V P_g^{-1})^2 : R_U \in G_D^0, R_V \in G_\Lambda^0 \} \right)^2 \right. \\ \left. + \left\| \log (D^{-1}(\pi_g \cdot \Lambda)) \right\|^2 \right\}. \quad (4.14)$$

Every minimal smooth scaling-rotation curve from X to Y corresponds to some minimal pair whose first element lies in the connected component $[(U, D)]$ of \mathcal{E}_X . A pair $((UR_U, D), (VR_V P_g^{-1}, \pi_g \cdot \Lambda))$ with $R_U \in G_D^0, R_V \in G_\Lambda^0$, and $g \in \tilde{S}_p^+$, is minimal if and only if the triple (g, R_U, R_V) is a minimizer of the expression in braces in (4.13).

Proof: From Proposition 2.5, we have

$$\mathcal{E}_X \times \mathcal{E}_Y = \left\{ (UR_U P_{g_1}^{-1}, \pi_{g_1} \cdot D), (VR_V P_{g_2}^{-1}, \pi_{g_2} \cdot \Lambda) : R_U \in G_D^0, R_V \in G_\Lambda^0; g_1, g_2 \in \tilde{S}_p^+ \right\}. \quad (4.15)$$

Using Proposition 4.3 and the fact the maps $g \mapsto \pi_g, g \mapsto P_g$, are homomorphisms, for all $g_1, g_2 \in \tilde{S}_p^+$ we have

$$d((UR_U P_{g_1}^{-1}, \pi_{g_1} \cdot D), (VR_V P_{g_2}^{-1}, \pi_{g_2} \cdot \Lambda)) \quad (4.16)$$

$$= d((UR_U, D), (VR_V (P_{g_1^{-1} g_2})^{-1}, \pi_{g_1^{-1} g_2} \cdot \Lambda)). \quad (4.17)$$

The map $(\tilde{U}, \tilde{D}) \mapsto (\tilde{U} P_{g_1}, \pi_{g_1}^{-1} \cdot \tilde{D})$ underlying the equality above is an isometry of $(SO \times \text{Diag}^+)(p)$ that preserves every fiber, hence carries a geodesic γ_1 with the endpoints in (4.16) into a geodesic γ_2 with the endpoints in (4.17) and that satisfies $F \circ \gamma_1 = F \circ \gamma_2$. Hence, every smooth scaling-rotation (SSR) curve from X to Y is of the form $F \circ \gamma$ where $\gamma : [0, 1] \rightarrow (SO \times \text{Diag}^+)(p)$ is a geodesic with $\gamma(0) = (UR_U, D) \in [(U, D)]$ and $\gamma(1) \in \mathcal{E}_Y$.

Suppose γ_1, γ_2 are two such geodesics, with $\gamma_i(1) = (VR_V P_{g_i}^{-1}, \pi_{g_i} \cdot \Lambda)$, $i = 1, 2$. If $g_2 = h_D g_1 h_\Lambda$, with $h_D \in \Gamma_{JD}^0$ and $h_\Lambda \in \Gamma_{JA}^0$, then

$$d((UR_U, D), (VR_V P_{g_2}^{-1}, \pi_{g_2} \cdot \Lambda)) \\ = d((UR_U, D), (VR_V h_\Lambda^{-1} P_{g_1}^{-1} h_D^{-1}, \pi_{h_D} \cdot \pi_{g_1} \cdot \pi_{h_\Lambda} \cdot \Lambda)) \\ = d((UR_U h_D, \pi_{h_D} \cdot D), (VR_V h_\Lambda^{-1} P_{g_1}^{-1}, \pi_{g_1} \cdot \pi_{h_\Lambda} \cdot \Lambda)) \\ = d((UR_{U,1}, D), (VR_{V,1} P_{g_1}^{-1}, \pi_{g_1} \cdot \Lambda))$$

where $R_{U,1} = R_U h_D \in G_D^0$ and $R_{V,1} = R_V h_\Lambda^{-1} \in G_\Lambda^0$. The same argument as in the preceding paragraph shows that the SSR curve determined by the pair $((UR_U, D), (VR_V P_{g_2}^{-1}, \pi_{g_2} \cdot \Lambda))$ is the same as the SSR curve determined by the pair $((UR_{U,1}, D), (VR_{V,1} P_{g_1}^{-1}, \pi_{g_1} \cdot \Lambda))$. Hence any representative $g \in \tilde{S}_p^+$ of a given $(\Gamma_{J_D}^0, \Gamma_{J_\Lambda}^0)$ double-coset determines the same set of SSR curves (the set of curves $F \circ \gamma$, where γ is a geodesic from (UR_U, D) to $(VR_V g^{-1}, \pi_g \cdot \Lambda)$ for some $R_U \in G_D^0, R_V \in G_\Lambda^0$) as does any other representative of that double-coset. The Proposition now follows. \blacksquare

Proposition 4.10 remains true if Z is replaced by \tilde{S}_p^+ in (4.13), or by a set of representatives of the left-coset space $S_p/\Gamma_{J_\Lambda}^0$ or the right-coset space $\Gamma_{J_D}^0 \backslash S_p$. However, taking Z as in the Proposition reduces the combinatorial complexity of the minimization problem when the eigenvalues of X or Y are not all distinct. In Sections 5 and 6 of [8], this proposition is applied to simplify the concrete computations of $d_{\mathcal{SR}}(X, Y)$ and $\mathcal{M}(X, Y)$, respectively, in the case $p = 3$.

4.3. Uniqueness questions for MSSR curves

As noted in Section 4.1, for all $X, Y \in \text{Sym}^+(p)$ there always exists an MSSR curve from X to Y , the projection of some F -minimal geodesic. *A priori*, different F -minimal geodesics could project to the same MSSR curve or to different MSSR curves. It is natural to ask: Under what conditions on (X, Y) is there a unique MSSR curve from X to Y ? When uniqueness fails, *how* does it fail, and what can we say about the set $\mathcal{M}(X, Y)$?

For uniqueness to fail for given X, Y , there must be distinct F -minimal geodesics $\gamma_i : [0, 1] \rightarrow M(p)$, whose endpoints are minimal pairs $((U_i, D_i), (V_i, \Lambda_i)) \in \mathcal{E}_X \times \mathcal{E}_Y$, $i = 1, 2$, such that $F \circ \gamma_1 \neq F \circ \gamma_2$. The “how” question above concerns the following two possibilities (not mutually exclusive):

1. “Type I non-uniqueness”: There exist such γ_i whose endpoints are *distinct* minimal pairs $((U_i, D_i), (V_i, \Lambda_i))$.
2. “Type II non-uniqueness”: There exist such γ_i whose endpoints are *the same* minimal pair $((U, D), (V, \Lambda))$.

Since for any $D, \Lambda \in \text{Diag}^+(p)$ the minimal geodesic from D to Λ is unique, Type II non-uniqueness with minimal pair $((U, D), (V, \Lambda))$ is equivalent to the existence of two or more minimal geodesics from U to V , which is equivalent to each of U, V being in the cut-locus (in $SO(p)$) of the other. It will be convenient for us to have some other terminology for this situation:

Definition 4.11. Call a pair of points (U, V) in $SO(p) \times SO(p)$ *geodesically antipodal* if one point is in the cut-locus of the other (equivalently, if each point is in the cut-locus of the other) and *geodesically non-antipodal* otherwise. Call a pair of points $((U, D), (V, \Lambda))$ in $M(p) \times M(p)$ geodesically antipodal if (U, V) is a geodesically antipodal pair in $SO(p) \times SO(p)$, and geodesically non-antipodal otherwise.

As mentioned earlier, the cut-locus of the identity $I \in SO(p)$ is precisely the set of all involutions in $SO(p)$. Furthermore, because of the invariance of the Riemannian metric g_{SO} , an element $V \in SO(p)$ is in the cut-locus of element U if and only if $V^{-1}U$ is in the cut-locus of I . Note that, for elements a, b of any group, ab is an involution $\iff ba$ is an involution $\iff b^{-1}a^{-1}$ is an involution $\iff a^{-1}b^{-1}$ is an involution. Thus, if any of the elements $V^{-1}U, UV^{-1}, U^{-1}V, VU^{-1}$ is an involution, so are all the others.

Note that a pair (U, V) in $SO(p)$ can be geodesically antipodal without either point being maximally remote from the other. (For example, with $p = 4$, the matrix $\text{diag}(-1, -1, 1, 1)$ is an involution, but is closer to the identity I than is the involution $-I$.) However, if (U, V) is geodesically antipodal, then there exists a (not necessarily unique) closed geodesic in $SO(p)$ containing U and V , isometric to a circle of some radius, such that U and V are antipodal points of this circle in the usual sense, and such that each of the two semicircles connecting U and V is a minimal geodesic between these two points.

One reason for our interest in Type II non-uniqueness is its effect on a true scaling-rotation metric ρ_{SR} on $\text{Sym}^+(p)$, mentioned earlier, that we construct from d_{SR} in [9]. Various constructions and assertions concerning this metric are simplified when we know that Type II non-uniqueness does not occur.

For small enough values of p , Type II non-uniqueness never occurs; for large enough p , it always occurs. Our proof of this requires a long digression from the topic of scaling-rotation distance and MSSR curves, so we defer the proof to Section 5. However, we will introduce here our main tool for ruling out Type II non-uniqueness, based on a property we call *sign-change reducibility* (for want of a better term) defined shortly.

To motivate the definition, let $X, Y \in \text{Sym}^+(p)$ and let $((U, D), (V, \Lambda)) \in \mathcal{E}_X \times \mathcal{E}_Y$ be a minimal pair. Then one minimizer (g, R_U, R_V) of the expression in brackets on the right-hand side of (4.13) is the triple (e, I, I) , where e is the identity element of \tilde{S}_p^+ . Hence for all $g \in \tilde{S}_p$ with $\pi_g \cdot \Lambda = \Lambda$ —i.e. for all $g \in K_{J_\Lambda}$ (see Notation 2.7 and equation (2.18))—we must have $d_{SO}(UP_g, V) = d_{SO}(U, VP_g^{-1}) \geq d_{SO}(U, V)$. But $\mathcal{I}_p^+ \subset K_J$ for all J (since sign-change matrices act trivially on $\text{Diag}^+(p)$), so, in particular, we must have $d_{SO}(UI_\sigma, V) \geq d_{SO}(U, V)$ for all $\sigma \in \mathcal{I}_p^+$.

Definition 4.12. Call a pair of points $(U, V) \in SO(p) \times SO(p)$ *sign-change reducible* if $d_{SO}(UI_\sigma, V) < d_{SO}(U, V)$ for some $\sigma \in \mathcal{I}_p^+$.

From the discussion preceding Definition 4.12, we have the following:

Corollary 4.13. Let $((U, D), (V, \Lambda)) \in M(p) \times M(p)$. If $(U, V) \in SO(p) \times SO(p)$ is sign-change reducible, then $((U, D), (V, \Lambda))$ is not a minimal pair.

■

Sign-change reducibility is studied in more detail in Sections 5–8. Below, we summarize some results proven there, and their consequences. Two of the main results are given in the following Proposition (proven in Section 8):

Proposition 4.14. (a) For $p \leq 4$, every geodesically antipodal pair (U, V) in $SO(p) \times SO(p)$ is sign-change reducible. (b) For $p \geq 11$, there exist geodesically antipodal pairs (U, V) in $SO(p) \times SO(p)$ that are not sign-change reducible.

Thus the largest dimension p_1 for which every geodesically antipodal pair (U, V) in $SO(p_1) \times SO(p_1)$ is sign-change reducible satisfies $4 \leq p_1 \leq 10$. A combination of theory and numerical evidence leads the authors to believe that p_1 is closer to 10 than to 4.

An immediate consequence of Proposition 4.14 (a) is the following. (Again, we do not believe the number “4” here is sharp.)

Corollary 4.15. For $p \leq 4$, every minimal pair in $M(p) \times M(p)$ is geodesically non-antipodal. Hence for $p \leq 4$, for all $X, Y \in \text{Sym}^+(p)$ for which $\mathcal{M}(X, Y) > 1$, the non-uniqueness is purely of Type I.

Part of the importance of sign-change reducibility comes from the following:

Proposition 4.16. Suppose that (U, V) is a pair in $SO(p) \times SO(p)$ that is not sign-change reducible. Then there exist $D, \Lambda \in \mathcal{D}_{\text{J}_{\text{top}}}$ such that the pair $((U, D), (V, \Lambda))$ is minimal.

We will prove this below. But first note that an immediate corollary of Propositions 4.14(b) and 4.16 is:

Corollary 4.17. For $p \geq 11$, there exist geodesically antipodal, minimal pairs $((U, D), (V, \Lambda)) \in \mathcal{S}_{\text{J}_{\text{top}}} \times \mathcal{S}_{\text{J}_{\text{top}}} \subset M(p) \times M(p)$. Hence, for $p \geq 11$, there exist $X, Y \in \mathcal{S}_{[\text{J}_{\text{top}}]} \subset \text{Sym}^+(p)$ for which the set $\mathcal{M}(X, Y)$ exhibits Type II non-uniqueness.

Thus sign-change reducibility is more than an *ad hoc* criterion for ruling out Type II non-uniqueness for small enough p . Proposition 4.16 and Corollary 4.17 show that, in some sense, sign-change reducibility is the *only* obstruction to having points X, Y in the top stratum of $\text{Sym}^+(p)$ for which $\mathcal{M}(X, Y)$ exhibits Type II nonuniqueness.

For X or Y not in the top stratum of $\text{Sym}^+(p)$, the relationship between Type II nonuniqueness and sign-change reducibility of minimal pairs in $\mathcal{E}_X \times \mathcal{E}_Y$ situation is more complicated to analyze. We do not investigate this relationship further in this paper.

To prove Proposition 4.16 we start with a lemma:

Lemma 4.18. Let $c > 0$. There exist $D, \Lambda \in \mathcal{D}_{\text{top}} := \mathcal{D}_{\text{J}_{\text{top}}}$ such that $\|\log(D^{-1}(\pi \cdot \Lambda))\|^2 > \|\log(D^{-1}\Lambda)\|^2 + c$ for all non-identity $\pi \in S_p$.

Proof: Let $c_1 = \sqrt{c/(3p)}$ and let $\{a_i\}_{i=1}^p$ be a sequence of numbers satisfying $a_{i+1} - a_i > (2\sqrt{p} + 1)c_1$ for $1 \leq i \leq p-1$. Then $|c + a_j - a_i| > 2\sqrt{p}c$ for all $i \neq j$. Let $D = \text{diag}(e^{a_1}, \dots, e^{a_p})$ and let $\Lambda = e^{c_1} D$. Then $D, \Lambda \in \mathcal{D}_{\text{top}}$ and $\|\log(D^{-1}\Lambda)\|^2 = \|c_1 I\|^2 = pc_1^2$.

Let $\pi \in S_p, \pi \neq \text{id}$, and let i be such that $\pi^{-1}(i) \neq i$. Then

$$\|\log(D^{-1}(\pi \cdot \Lambda))\|^2 \geq |c_1 + a_{\pi^{-1}(i)} - a_i|^2 > (2\sqrt{p}c_1)^2 = \|\log(D^{-1}\Lambda)\|^2 + c.$$

■

Proof of Proposition 4.16. Let $D, \Lambda \in \mathcal{D}_{\text{top}}$ such that

$$\|\log(D^{-1}(\pi \cdot \Lambda))\|^2 > \|\log(D^{-1}\Lambda)\|^2 + k \text{diam}(SO(p))^2 \quad (4.18)$$

for all non-identity $\pi \in S_p$; such D, Λ exist by Lemma 4.18. Let $X = F(U, D), Y = F(V, \Lambda)$. The subgroups G_D^0, G_Λ^0 of $SO(p)$ are trivial, as are the subgroups Γ_{JD}^0 and $\Gamma_{J\Lambda}^0$ of \tilde{S}_p^+ . Hence in Proposition 4.10 we have $Z = \tilde{S}_p^+$ and

$$\begin{aligned} d_{\mathcal{SR}}(X, Y)^2 &= \min_{g \in \tilde{S}_p^+} \left\{ k d_{SO}(U, VP_g^{-1})^2 + \|\log(D^{-1}(\pi_g \cdot \Lambda))\|^2 \right\} \\ &= \min_{\pi \in S_p} \left\{ k \min \left\{ d_{SO}(U, VP_g^{-1})^2 : g \in \tilde{S}_p^+, \pi_g = \pi \right\} \right. \\ &\quad \left. + \|\log(D^{-1}(\pi \cdot \Lambda))\|^2 \right\}. \end{aligned}$$

For all non-identity $\pi \in S_p$ and all $g_1, g_2 \in \tilde{S}_p^+$ with $\pi_{g_1} = \text{id}$. and $\pi_{g_2} = \pi$, using (4.18) we then have

$$\begin{aligned} d_M((U, D), (VP_{g_1}^{-1}, \pi_{g_1} \cdot \Lambda))^2 &= k d_{SO}(U, VP_{g_1}^{-1})^2 + \|\log(D^{-1}\Lambda)\|^2 \\ &\leq k \text{diam}(SO(p))^2 + \|\log(D^{-1}\Lambda)\|^2 \\ &< \|\log(D^{-1}(\pi \cdot \Lambda))\|^2 \\ &\leq d_M((U, D), (VP_{g_2}^{-1}, \pi_{g_2} \cdot \Lambda))^2. \end{aligned}$$

Hence the identity permutaton is the only element of S_p for which the expression inside the outer braces in (4.19) achieves the minimum over all $\pi \in S_p$. But $\{g \in \tilde{S}_p^+ : \pi_g = \text{id}\}$ is precisely the sign-change subgroup \mathcal{I}_p^+ , and by hypothesis (U, V) is not sign-change reducible. Hence

$$\begin{aligned} d_{\mathcal{SR}}(X, Y)^2 &= \min_{\sigma \in \mathcal{I}_p^+} \left\{ k d_{SO}(U, VI_\sigma)^2 + \|\log(D^{-1}\Lambda)\|^2 \right\} \\ &= k d_{SO}(U, V)^2 + \|\log(D^{-1}\Lambda)\|^2 \\ &= d_M((U, D), (V, \Lambda))^2. \end{aligned}$$

Thus $((U, D), (V, \Lambda))$ is a minimal pair. ■

4.4. When do two minimal pairs determine the same MSSR curve?

Proposition 4.10 is a starting-point for understanding the set $\mathcal{M}(X, Y)$ for all p and all $X, Y \in \text{Sym}^+(p)$: it assures us that, for any $(U, D) \in \mathcal{E}_X$, every MSSR curve from X to Y corresponds to some minimal pair whose first element lies in the connected component $[(U, D)]$ of \mathcal{E}_X . But even once we know all the minimal pairs, to completely understand $\mathcal{M}(X, Y)$ —or even just determine its cardinality—we need a way to tell whether MSSR curves corresponding to two (not necessarily distinct) minimal pairs with first point in $[(U, D)]$ are the same. (This is true whether the non-uniqueness, if any, in $\mathcal{M}(X, Y)$ is of Type I, Type II, or a mixture of both). Proposition 4.19 below provides such a tool. For $p = 3$, this proposition suffices for the computation of $\mathcal{M}(X, Y)$ in the “nontrivial” cases (those in which neither X nor Y lies in the bottom stratum of $\text{Sym}^+(3)$, at least one lies in the middle stratum—the unique stratum of $\text{Sym}^+(3)$ that is neither the top nor bottom stratum); see [8, Section 6].

Proposition 4.19. *Let $X, Y \in \text{Sym}^+(p)$, $X \neq Y$. For $i = 1, 2$ assume that $\chi_i = F \circ \gamma_i$ is a minimal smooth scaling-rotation curve from X to Y corresponding to the minimal pair $((UR_{U,i}, D), (VR_{V,i}P_{g_i}^{-1}, \Lambda_i))$, where $R_{U,i} \in G_D^0$, $R_{V,i} \in G_\Lambda^0$, $g_i \in \tilde{S}_p^+$, $\Lambda_i = \pi_{g_i} \cdot \Lambda$, and $\gamma_i : [0, 1] \rightarrow M(p)$ is a geodesic. (We do not assume that the two minimal pairs are distinct.) Then $\chi_1 = \chi_2$ if and only if the following two conditions hold.*

(i) *Both pairs $(UR_{U,i}, VR_{V,i}P_{g_i}^{-1})$ are geodesically non-antipodal and*

$$R_{V,2}P_{g_2}^{-1}R_{U,2}^{-1} = R_{V,1}P_{g_1}^{-1}R_{U,1}^{-1}, \quad (4.19)$$

or both pairs are geodesically antipodal and

$$(\text{proj}_{SO(p)}\gamma_1'(0))R_{U,1}^{-1} = (\text{proj}_{SO(p)}\gamma_2'(0))R_{U,2}^{-1}, \quad (4.20)$$

where for any $(U', D') \in M(p)$, $\text{proj}_{SO(p)}$ denotes the natural projection $T_{(U', D')}M(p) \rightarrow T_{U'}SO(p)$.

(ii) *There exist $g \in \tilde{S}_p^+$, $R \in G_{D, \Lambda_1}^0$ such that*

$$D = \pi_g \cdot D, \quad (4.21)$$

$$\Lambda_2 = \pi_g \cdot \Lambda_1, \quad (4.22)$$

$$\text{and } R_{U,1}^{-1}R_{U,2} = RP_g^{-1}. \quad (4.23)$$

Equation (4.20) implies equation (4.19), so (4.19) is always a necessary condition for the equality $\chi_1 = \chi_2$.

In Proposition 4.19, in the geodesically non-antipodal case we use *endpoint data* to tell whether the projections to $\text{Sym}^+(p)$ of two minimal geodesics from \mathcal{E}_X to \mathcal{E}_Y are equal. We will deduce this proposition from the following theorem, proven in [11], that gives a criterion based on *initial-value data* to tell whether the projections of two geodesics emanating from \mathcal{E}_X are equal. In this theorem, $G_{D,L} := G_D \cap G_L$, $\mathfrak{g}_{D,L} := \mathfrak{g}_D \cap \mathfrak{g}_L$ (the Lie algebra of $G_{D,L}$), and for $A \in SO(p)$, $\text{ad}_A : \mathfrak{so}(p) \rightarrow \mathfrak{so}(p)$ is the linear map defined by $\text{ad}_A(B) = [A, B]$.

Theorem 4.20 ([11, Theorem 3.8]). *For $i = 1, 2$ let $(U_i, D_i) \in M(p)$, $A_i \in \mathfrak{so}(p)$, $L_i \in \text{Diag}(p)$, and let $\check{A}_i = U_i^{-1} A_i U_i$. Let I be a positive-length interval containing 0. Then the smooth scaling-rotation curves $\chi_i := \chi_{U_i, D_i, A_i, L_i} : I \rightarrow \text{Sym}^+(p)$ are identical if and only if (i) $\check{A}_2 - \check{A}_1 \in \mathfrak{g}_{D_1, L_1}$, (ii) $(\text{ad}_{\check{A}_2})^j(\check{A}_1) \in \mathfrak{g}_{D_1, L_1}$ for all $j \geq 1$, and (iii) there exist $R \in G_{D_1, L_1}$ and $g \in \tilde{S}_p^+$, such that $U_2 = U_1 R P_g^{-1}$, $D_2 = \pi_g \cdot D_1$, and $L_2 = \pi_g \cdot L_1$.¹*

To deduce Proposition 4.19 from Theorem 4.20, we first prove two lemmas. Beyond helping us to prove the Proposition, these lemmas may be useful in further analysis of MSSR curves. In these lemmas, for any $X \in \text{Sym}^+(p)$ we write \mathfrak{g}_X for the Lie algebra of the stabilizer $G_X := \{U \in G : UXU^T = X\}$; thus $\mathfrak{g}_X = \{A \in \mathfrak{so}(p) : AX = XA\}$. (Observe that this is consistent with the notation G_D, \mathfrak{g}_D introduced earlier for diagonal matrices.)

Lemma 4.21. *Let $X, Y \in \text{Sym}^+(p)$ and suppose that $\chi : [0, 1] \rightarrow \text{Sym}^+(p)$ is a minimal smooth rotation-scaling curve with $X := \chi(0) \neq Y := \chi(1)$. Let $\gamma = \gamma_{U, D, A, L} : [0, 1] \rightarrow SO(p) \times \text{Diag}^+(p)$ be a geodesic for which $\chi = F \circ \gamma$. Then $A \in (\mathfrak{g}_X)^\perp \cap (\mathfrak{g}_Y)^\perp$, where the orthogonal complements are taken in $\mathfrak{so}(p)$.*

Proof: Since γ is a smooth curve of minimal length connecting the submanifolds \mathcal{E}_X and \mathcal{E}_Y of $(SO \times \text{Diag})^+(p)$, the velocity vectors $\gamma'(0), \gamma'(1)$ must be perpendicular to the tangent spaces $T_{\gamma(0)}\mathcal{E}_X, T_{\gamma(1)}\mathcal{E}_Y$, respectively ([3, Proposition 1.5]). Making natural tangent-space identifications, we have $T_{\gamma(0)}\mathcal{E}_X = T_{(U, D)}\mathcal{E}_X = U\mathfrak{g}_D \oplus \{0\} \subset U\mathfrak{g}_D \oplus \text{Diag}(p)$, where $U\mathfrak{g}_D := \{UC : C \in \mathfrak{g}_D\}$. Let $\check{A} = U^{-1}AU$. Since $\gamma'(0) = (U\check{A}, DL)$, and the Riemannian metric we are using on $SO(p)$ is left-invariant, the condition $\gamma'(0) \perp T_{\gamma(0)}\mathcal{E}_X$ is equivalent to $\check{A} \in (\mathfrak{g}_D)^\perp$, hence to $A \in U(\mathfrak{g}_D)^\perp U^{-1}$. Using additionally the right-invariance of the metric on $\mathfrak{so}(p)$, we have $U(\mathfrak{g}_D)^\perp U^{-1} = (U\mathfrak{g}_D U^{-1})^\perp$. From general group-action properties, it is easily seen that $U\mathfrak{g}_D U^{-1} = \mathfrak{g}_{UDU^{-1}}$. Since $UDU^{-1} = X$, it follows that $A \in (\mathfrak{g}_X)^\perp$. A similar argument at the point $(V, \Lambda) := \gamma(1)$ shows that $A \in (\mathfrak{g}_{V\Lambda V^{-1}})^\perp = (\mathfrak{g}_Y)^\perp$. ■

¹In [11, Theorem 3.8], g was actually required to be of the form $(s(\pi), \pi)$, where s is as in (??). But the same argument as in the proof of Proposition 2.5 shows that this restriction can be removed.

Lemma 4.22. *In the setting of Theorem 4.20, assume that the smooth scaling-rotation curve χ_1 is minimal. Then conditions (i) and (ii) in the theorem can be replaced by the single condition $A_2 = A_1$.*

Proof: With notation as in Theorem 4.20, assume that $\chi_2 = \chi_1$. Then the Theorem implies that $U^{-1}(A_2 - A_1)U \in \mathfrak{g}_{D,L} \subset \mathfrak{g}_D$, implying that $A_2 - A_1 \in U\mathfrak{g}_D U^{-1} = \mathfrak{g}_X$ (as in the proof of Lemma 4.21). But since χ_1 is minimal, Lemma 4.21 implies that both A_2 and A_1 lie in $(\mathfrak{g}_X)^\perp$, hence that $A_2 - A_1 \in (\mathfrak{g}_X)^\perp$. Hence $A_2 - A_1 = 0$, i.e. $A_2 = A_1$.

Conversely, assume that $A_2 = A_1$. Then conditions (i) and (ii) are satisfied trivially. \blacksquare

Proof of Proposition 4.19: For $i \in \{1, 2\}$ let $U_i = UR_{U,i}$, $V_i = VR_{V,i}P_{g_i}^{-1}$, and $\Lambda_i = \pi_{g_i} \cdot \Lambda_i$.

By hypothesis $\chi_i = F \circ \gamma_i$, where $\gamma_i = \gamma_{U_i,D,A_i,L_i} : [0, 1] \rightarrow M(p)$ (for some $A_i \in \mathfrak{so}(p)$, $L_i \in \text{Diag}(p)$) is a minimal geodesic from (U_i, D) to (V_i, Λ_i) . Hence $L_i = \log(\Lambda_i D^{-1})$ and $A_i \in \log(V_i U_i^{-1})$ (we write “ \in ” rather than “ $=$ ” since if R is an involution, “ $\log R$ ”, as we have defined it, is a set with more than one element; see Section 4.1).

It is straightforward to show that $G_{D,L_i} = G_{D,\Lambda_i}$. From Lemma 4.22, the conditions (i) and (ii) in Theorem 4.20 in the equality-conditions for χ_1 and χ_2 can be replaced by the single condition $A_2 = A_1$.

If $A_2 = A_1$ then $V_2 U_2^{-1} = V_1 U_1^{-1}$, implying that either both pairs $\{(U_i, V_i)\}$ are geodesically antipodal or both are geodesically non-antipodal. In the converse direction, suppose that the pairs $\{(U_i, V_i)\}$ are geodesically non-antipodal and that $V_2 U_2^{-1} = V_1 U_1^{-1}$. Then $A_2 = \log(V_2 U_2^{-1}) = \log(V_1 U_1^{-1}) = A_1$. Whether or not the pairs $\{U_i, V_i\}$ are geodesically antipodal, by definition $(\text{proj}_{SO(p)} \gamma'_i(0))U_i^{-1} = A_i$, so if (4.20) holds then $A_2 = A_1$. Hence the condition $A_2 = A_1$ is equivalent to condition (i) in Proposition 4.19.

Next, letting D play the role of D_1 in Theorem 4.20, condition (iii) in the Theorem is equivalent to the existence of $g \in \tilde{S}_p^+$, $R \in G_{D,\Lambda_1}$ such that $D = \pi_g \cdot D$, $L_2 = \pi_g \cdot L_1$, and $U_2 = U_1 R P_g^{-1}$. But for all such R, π , we have $R P_g^{-1} = R_0 P_{g_0}^{-1}$ for some $R_0 \in G_{D,\Lambda_1}^0$ and $g_0 \in \tilde{S}_p^+$ with $\pi_{g_0} = \pi_g$. Furthermore, for any $\pi \in S_p$, if $\pi \cdot D = D$ then $L_2 = \pi \cdot L_1 \iff \Lambda_2 = \pi \cdot \Lambda_1$. Hence, under the hypotheses of Proposition 4.19, condition (iii) in Theorem 4.20 is equivalent to condition (ii) stated in the Proposition.

This establishes the “if and only if” statement in the Proposition. The final statement of the proposition follows from the fact that, in the notation of this proof, (4.20) is the equality $A_2 = A_1$ (after multiplying both sides of (4.20) on the right by U^{-1}), an equality that implies $V_2 U_2^{-1} = \exp(A_2) = \exp(A_1) = V_1 U_1^{-1}$. \blacksquare

5. Involutions, sign-change reducibility, and distance between subspaces of \mathbf{R}^p

In this section we begin our study of sign-change reducibility. This culminates in Section 8 with the proof of Proposition 4.14 (which, as we have seen, implies Corollary 4.17, our main result concerning Type II nonuniqueness), but we discover some other interesting facts along the way. As we shall see, questions concerning the seemingly *ad hoc* notion of sign-change reducibility can be translated into questions about distances between subspaces of \mathbf{R}^p ; for example, Proposition 5.11 states the equivalence between a sign-change-reducibility question and a question purely about the geometry of the Grassmannian $\text{Gr}_m(\mathbf{R}^p)$ (endowed with a standard metric). Thus, some unexpected benefits of our investigation of Type II nonuniqueness are results, possibly of independent interest, concerning the geometry of Grassmannians and, more generally, principal angles between subspaces of \mathbf{R}^p .

Recall from Section 4.3 that two points $U, V \in SO(p)$ are geodesically antipodal if and only if $V^{-1}U$ is an involution. Since $d_{SO}(U, V) = d_{SO}(V^{-1}U, I)$, the set of distances between geodesically antipodal points in $SO(p)$ is the same as the set of distances between the identity and involutions. Thus to understand which (if any) geodesically antipodal pairs (U, V) in $SO(p)$ are sign-change reducible, it suffices to study the case $(U, V) = (R, I)$, where R is an involution.

Definition 5.1.

1. Call $R \in SO(p)$ *sign-change reducible* if $d_{SO}(RI_\sigma, I) < d_{SO}(R, I)$ for some $\sigma \in \mathcal{I}_p^+$ (equivalently, if the pair (R, I) is sign-change reducible). Note that sign-change reducibility of the pair (U, V) , as previously defined in Definition 4.12, is equivalent to sign-change reducibility of $V^{-1}U$.
2. For $\sigma = (\sigma_1, \dots, \sigma_p) \in \mathcal{I}_p$, define the *level* of σ , written $\text{level}(\sigma)$, to be $\#\{i : \sigma_i = -1\}$.
3. For any involution $R \in SO(p)$, define the *level* of R , written $\text{level}(R)$, to be $\dim(E_{-1}(R))$. We write $\text{Inv}(p)$ for the set of involutions in $SO(p)$, and for $0 < m \leq p$ we write $\text{Inv}_m(p)$ for the set of involutions in $SO(p)$ of level m . Note that $\dim(E_{-1}(R))$ is even for any $R \in SO(p)$, so $\text{Inv}_m(p)$ is empty unless m is even and at least 2. Thus $\text{Inv}(p) = \bigcup_{\text{even } m \geq 2} \text{Inv}_m(p)$ (a disjoint union).
4. Let $R \in SO(p)$ be an involution. We say that R is *reducible by a sign-change of level m* if there exists $\sigma \in \mathcal{I}_p^+$ of level m such that $d_{SO}(RI_\sigma, I) < d_{SO}(R, I)$.

Observe that for non-identity $\sigma \in \mathcal{I}_p^+$, the matrix I_σ is an involution in $SO(p)$, and $\text{level}(\sigma) = \text{level}(I_\sigma)$.

Remark 5.2 (Involutions and Grassmannians). The space $\text{Inv}(p)$ can be naturally identified with a disjoint union of Grassmannians, because an involution $R \in SO(p)$ is completely determined by its (-1) -eigenspace $E_{-1}(R)$. Let $\text{Gr}_m(\mathbf{R}^p)$ denote the Grassmannian of m -planes in \mathbf{R}^p , and for even $m \in (0, p]$

define $\Phi_{m,p} : \text{Gr}_m(\mathbf{R}^p) \rightarrow \text{Inv}_m(p)$ to be the map carrying $W \in \text{Gr}_m(\mathbf{R}^p)$ to the involution in $SO(p)$ whose (-1) -eigenspace is W . (Thus $E_{-1}(R) = \Phi_{m,p}^{-1}(R)$ for all $R \in \text{Inv}_m(p)$.) Concretely, letting $\pi_V : \mathbf{R}^p \rightarrow V$ denote orthogonal projection onto any subspace V , and letting P_V denote the matrix of π_V with respect to the standard basis of \mathbf{R}^p , the map $\Phi_{m,p}$ is given by

$$\Phi_{m,p}(W) = P_{W^\perp} - P_W = I - 2P_W, \quad (5.1)$$

reflection about the $(p-m)$ -plane W^\perp . It is not hard to show that $\text{Inv}_m(p)$ is a submanifold of $SO(p)$ and that $\Phi_{m,p}$ is a diffeomorphism from $\text{Gr}_m(\mathbf{R}^p)$ to this submanifold.

Our study of sign-change reducibility of involutions will make frequent use of the *normal form* of an element of $SO(p)$, so we review this before proceeding.

5.1. Normal form and distance to the identity in $SO(p)$

Let $k = \lfloor \frac{p}{2} \rfloor$. Recall that every $R \in SO(p)$ has a *normal form*: a block-diagonal matrix that, for p even, is of the form

$$R(\theta_1, \dots, \theta_k) = \begin{bmatrix} C(\theta_1) & & & \\ & C(\theta_2) & & \\ & & \ddots & \\ & & & C(\theta_k) \end{bmatrix}, \quad (5.2)$$

where

$$C(\theta) = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix} \quad (5.3)$$

and where $\theta_i \in [0, \pi]$, $1 \leq i \leq k$. (This can be derived quickly from the normal form of an antisymmetric matrix, since the compactness of $SO(p)$ guarantees that the exponential map $\mathfrak{so}(p) \rightarrow SO(p)$ is onto.) For the odd- p case, the normal-form matrix is the matrix (5.2) with one more row and column appended, and with a 1 in the lower right-hand corner (and zeroes everywhere else in the last row and column). In this case we define $\theta_{k+1} = 0$, so that for both even and odd p we can use the notation $R(\theta_1, \dots, \theta_{\lceil p/2 \rceil})$ for the normal form.

Note that

$$C(\theta) = \exp(\theta J) \quad \text{where} \quad J = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}. \quad (5.4)$$

For each $R \in SO(p)$ there exists an orthonormal basis of \mathbf{R}^p with respect to which the linear transformation $\mathbf{R}^p \rightarrow \mathbf{R}^p$, $v \mapsto Rv$, has matrix $R(\theta_1, \dots, \theta_{\lceil p/2 \rceil})$. Thus there exists $Q \in O(p)$ such that

$$R = QR(\theta_1, \dots, \theta_{\lceil p/2 \rceil})Q^{-1}. \quad (5.5)$$

The normal form of a given R is unique up to ordering of the blocks; the multi-set $\{\theta_1, \dots, \theta_{\lceil p/2 \rceil}\}$ is uniquely determined by R . From (5.3) and (5.5) we have

$$R = Q \exp(A(\theta_1, \dots, \theta_{\lceil p/2 \rceil})) Q^{-1} = \exp(QA(\theta_1, \dots, \theta_{\lceil p/2 \rceil})Q^{-1}) \quad (5.6)$$

where $A(\theta_1, \dots, \theta_{\lceil p/2 \rceil})$ is the block-diagonal matrix obtained by replacing $C(\theta_i)$ by $\theta_i J$ in (5.2), $1 \leq i \leq \lfloor p/2 \rfloor$, and, in the odd- p case, replacing the 1 in the lower right-hand corner by 0. Since the normal form is unique up to block-ordering, it follows that

$$d_{SO}(R, I)^2 = \sum_{i=1}^{\lfloor p/2 \rfloor} \theta_i^2 = \sum_{i=1}^{\lceil p/2 \rceil} \theta_i^2 \quad (5.7)$$

Furthermore, from (5.5) and (5.3) it follows that

$$R_{\text{sym}} := \frac{R + R^T}{2} = Q \begin{bmatrix} \cos \theta_1 & I_{2 \times 2} & & & \\ & \cos \theta_2 & I_{2 \times 2} & & \\ & & & \ddots & \\ & & & & \cos \theta_k & I_{2 \times 2} \end{bmatrix} Q^{-1} \quad (5.8)$$

if p is even; for odd p we again just append one more row and column of the middle matrix, with a 1 in the lower right-hand corner. Hence the values $\cos \theta_i$ (and therefore the values $\theta_i \in [0, \pi]$) can be recovered from R as the eigenvalues of R_{sym} , with the multiplicity of an eigenvalue λ of R_{sym} equal to twice the multiplicity m_λ of λ in the list $\cos \theta_1, \dots, \cos \theta_k$ in the even- p case; for odd p the only difference is that multiplicity of the eigenvalue 1 of R_{sym} is $2m_1 + 1$.

Remark 5.3 (Normal form, involutions, and distances to identity).

Writing $R \in SO(p)$ in the form (5.5), it is easily seen that R is an involution if and only if (i) for each i , θ_i is either 0 or π , and (ii) $\theta_i = \pi$ for at least one i . For such R , if $\theta_i = \pi$ for exactly m values of i , then $\|A(\theta_1, \dots, \theta_{\lceil p/2 \rceil})\|^2 = m\pi^2$. Hence if $R \in SO(p)$ is an involution of level m , then

$$d_{SO}(R, I)^2 = \frac{m}{2} \pi^2. \quad (5.9)$$

Thus

$$\{d_{SO}(R, I) : R \in SO(p), R \text{ an involution}\} = \left\{ \sqrt{m} \pi : 1 \leq m \leq \left\lfloor \frac{p}{2} \right\rfloor \right\}. \quad (5.10)$$

Using (5.6) it can also be shown that for every non-involution $R \in SO(p)$, there is a unique $A \in \mathfrak{so}(p)$ of smallest norm such that $\exp(A) = R$.

Notation 5.4.

1. Given $R \in SO(p)$ and angles $\theta_1, \dots, \theta_{\lfloor p/2 \rfloor} \in [0, \pi]$ for which $R(\theta_1, \dots, \theta_{\lfloor p/2 \rfloor})$ is a normal form of R , we define “redundant normal-form angles” $\tilde{\theta}_i \in [0, \pi]$, $1 \leq i \leq p$, by

$$\tilde{\theta}_{2i-1} = \tilde{\theta}_{2i} = \theta_i, 1 \leq i \leq k = \left\lfloor \frac{p}{2} \right\rfloor; \quad \tilde{\theta}_p = 0 \text{ if } p = 2k + 1. \quad (5.11)$$

2. For any square matrix A we write $E_\lambda(A)$ for the λ -eigenspace of A .

Note that (5.7) can now be written as

$$d_{SO}(R, I)^2 = \frac{1}{2} \sum_{i=1}^p \tilde{\theta}_i^2. \quad (5.12)$$

5.2. Sign-change reducibility, distances in Grassmannians, and a half-angle relation

In this section we state and discuss several results, but defer their proofs to later sections.

For $p \leq 4$ one can show, without appealing to Proposition 5.6 below, that every involution in $SO(p)$ is sign-change reducible. (This sign-change reducibility holds for trivial reasons for when $p = 2$; holds for slightly less trivial reasons, mentioned later in Remark 5.12, for $p = 3$; and can be shown to hold for $p = 4$ using a quaternionic approach.) It is reasonable to wonder whether this holds for *all* p :

Question 5.5. *Let $p \geq 2$. Is every involution in $SO(p)$ sign-change reducible?*

Our motivation for this question is not just generalization for its own sake, however. Potential Type II non-uniqueness complicates several aspects of the analysis of scaling-rotation distance and the associated metric ρ_{SR} studied in [9]. To understand whether the “Type II non-uniqueness” defined in Section 4.3 can occur, we need to know whether a geodesically antipodal pair in $M(p)$ can be minimal. (As discussed in Section 4.3, a geodesically non-antipodal minimal pair in $M(p)$ uniquely determines an MSSR curve in $\text{Sym}^+(p)$.) A sufficient condition for any pair $((U, D), (V, \Lambda))$ in $M(p) \times M(p)$ to be *non*-minimal is that the pair $(U, V) \in SO(p)$ be sign-change reducible. Since sign-change reducibility of involutions rules out the possibility of Type II non-uniqueness, and all involutions are sign-change reducible for $p \leq 4$, it is natural to ask Question 5.5 and wish for the answer to be yes.

The answer, however, is more complicated. We shall see that the answer to Question 5.5 is yes for $p \leq 4$ and no for $p \geq 11$ (we do not know the answer for $5 \leq p \leq 10$), but that for all p , involutions of high enough level are sign-change reducible—moreover, by a sign-change of the same level:

Proposition 5.6. *Let $R \in SO(p)$ be an involution for which $\text{level}(R) \geq \frac{1}{2}p$. Then there exists $\sigma \in \mathcal{I}_p^+$, with $\text{level}(\sigma) = \text{level}(R)$, such that $d_{SO}(RI_\sigma, I) < d_{SO}(R, I)$.*

We defer the proof to Section 7.

Since $\text{level}(R) = \dim(E_{-1}(R)) \geq 2$ for every involution R , Proposition 5.6 (once proved) immediately establishes Proposition 4.14(a) and Corollary 4.15: for $p \leq 4$, all involutions are sign-change reducible, and hence all minimal pairs in $M(p) \times M(p)$ are geodesically non-antipodal.

We shall see below (Proposition 5.11) that sign-change reducibility by a sign-change of the same level is equivalent to a statement purely about the geometry of Grassmannians. For reasons given shortly, it seems likely to the authors that the “same level” condition appearing in Proposition 5.6 is optimal (even without the “ $\text{level}(R) \geq \frac{1}{2}p$ ” restriction) in the sense that $\min_{\sigma \in \mathcal{I}_p^+} \{d_{SO}(RI_\sigma, I)\}$ is achieved by a sign-change matrix $\sigma \in \mathcal{I}_p^+$ for which $\text{level}(\sigma) = \text{level}(R)$. If this is true, then the analysis of whether an involution R is sign-change reducible simplifies; we need only consider $\sigma \in \mathcal{I}_p^+$ of the same level as R . This (potential) simplification is actually of greater value to us than knowing, for a given $R \in \text{Inv}(p)$, whether all minimizers of $d_{SO}(RI_\sigma, I)$ have the same level as R , so we state only the following weaker conjecture:

Conjecture 5.7. *Let $m \geq 2$ be even, and let $R \in SO(p)$ be an involution of level m . If R is sign-change reducible, then it is reducible by a sign-change of level m .*

In Section 7 we will prove the following special case of this conjecture:

Proposition 5.8. *Conjecture 5.7 is true for $m = 2$.*

The reason we expect more generally that $\min_{\sigma \in \mathcal{I}_p^+} \{d_{SO}(RI_\sigma, I)\}$ is achieved by a σ for which $\text{level}(\sigma) = \text{level}(R)$ is as follows. Every sign-change matrix $I_{\sigma_1} \in \mathcal{I}_p^+$ is itself an involution, and satisfies

$$d_{SO}(I_{\sigma_1}I_{\sigma_1}, I) = 0 < \min\{d_{SO}(I_{\sigma_1}I_\sigma, I) : \sigma \in \mathcal{I}_p^+, \sigma \neq \sigma_1\}.$$

Thus for $R \in SO(p)$ sufficiently close to I_{σ_1} , we have

$$d_{SO}(RI_{\sigma_1}, I) < \min\{d_{SO}(RI_\sigma, I) : \sigma \in \mathcal{I}_p^+, \sigma \neq \sigma_1\}.$$

The function carrying an involution in $R \in SO(p)$ to $\text{level}(R)$ is continuous, so for $R \in \text{Inv}(p)$ sufficiently close to I_{σ_1} we also have $\text{level}(R) = \text{level}(\sigma_1)$. Hence for every $R \in \text{Inv}(p)$ sufficiently close to a sign-change matrix, $\min_{\sigma \in \mathcal{I}_p^+} \{d_{SO}(RI_\sigma, I)\}$ is achieved by a sign-change matrix having the same level as R . It seems plausible that this remains true even without the “sufficiently close to a sign-change matrix” restriction.

As noted in Remark 5.2, for even $m \geq 2$ the space $\text{Inv}_m(p)$ is diffeomorphic to the Grassmannian $\text{Gr}_m(\mathbf{R}^p)$. This Grassmannian carries a Riemannian metric

induced by Riemannian submersion from $(SO(p), g_{SO})$. It is known that the associated squared geodesic-distance between two points $W, Z \in \text{Gr}_m(\mathbf{R}^p)$ is, up to a constant factor, simply the sum of squares of the principal angles between the two m -planes W, Z .² Choosing the normalization in which the squared geodesic distance $d_{Gr}(W, Z)^2$ equals the sum of squares of the principal angles (equation (6.1) below), we will prove the following in Section 6:

Proposition 5.9. *The map $\Phi = \Phi_{m,p} : (\text{Gr}_m(\mathbf{R}^p), d_{Gr}) \rightarrow (\text{Inv}_m(p), d_{SO})$ (see (5.1)) is an isometry, up to a constant factor of 2:*

$$d_{SO}(\Phi(W), \Phi(V)) = 2d_{Gr}(W, V) \quad (5.13)$$

for all $W, V \in \text{Gr}_m(\mathbf{R}^p)$.

We derive Proposition 5.9 from a general half-angle relation proven in Section 6:

Proposition 5.10. *Let R_1, R_2 be involutions in $SO(p)$. For $i = 1, 2$ let $m_i = \dim(E_{-1}(R_i))$, and let $m = \min\{m_1, m_2\}$. Let $\{\theta_i \in [0, \pi]\}_{i=1}^{\lceil p/2 \rceil}$ be angles for which $R(\theta_1, \dots, \theta_{\lceil p/2 \rceil})$ is a normal form of the product $R_1 R_2$, and let $\{\tilde{\theta}_i\}_{i=1}^p$ be as defined in (5.11). Then for some injective map $\iota : \{1, 2, \dots, m\} \rightarrow \{1, 2, \dots, p\}$, the principal angles between $E_{-1}(R_1)$ and $E_{-1}(R_2)$ satisfy*

$$\phi_j(E_{-1}(R_1), E_{-1}(R_2)) = \frac{\tilde{\theta}_{\iota(j)}}{2}, \quad 1 \leq j \leq m. \quad (5.14)$$

For every $i \notin \text{range}(\iota)$, the angle $\tilde{\theta}_i$ is either 0 or π .

In other words, as stated in the introduction: for any two involutions $R_1, R_2 \in SO(p)$, each of the principal angles between $E_{-1}(R_1)$ and $E_{-1}(R_2)$ is exactly half a correspondingly indexed normal-form angle of $R_1 R_2$.

Proposition 5.9 can also be proven by purely Riemannian methods, but the proof we give, via Proposition 5.10, is independent in the sense that it does not make any use of a Riemannian metric on $\text{Gr}_m(\mathbf{R}^p)$; see Remark 6.5.

In Section 6, after proving Proposition 5.9 we will use it to deduce the following:

Proposition 5.11. *Let m, p be integers with m even and $0 < m \leq p$. Then the following two statements are equivalent:*

1. *Every involution $R \in SO(p)$ of level m is sign-change reducible by a sign-change of level m .*

²This fact follows from Wong's results on geodesics in [15], and has been cited elsewhere in the literature (e.g. [5, p. 337]), though the explicit statement does not appear in [15].

2. For every $W \in \text{Gr}_m(\mathbf{R}^p)$, there exists a coordinate m -plane \mathbf{R}^J (see Notation 6.1) such that

$$d_{Gr}(W, \mathbf{R}^J)^2 < \frac{m\pi^2}{8}. \quad (5.15)$$

In other words, the sign-change reducibility asserted in Statement 1 of the Proposition is equivalent to a statement purely about the geometry of Grassmannians (with the metric d_{Gr}), namely that the coordinate m -planes in \mathbf{R}^p form a “lattice” of $\binom{p}{m}$ points in $\text{Gr}_m(\mathbf{R}^p)$ such that every point in $\text{Gr}_m(\mathbf{R}^p)$ is within distance $(m\pi^2/8)^{1/2}$ of some lattice-point. This gives us a geometric way to tackle Question 5.5, at least for sign-change reducibility of an involution R by a sign-change matrix of the same level. However, the authors do not know a formula for $\min_{J \in \mathcal{J}_m} \{d_{Gr}(W, \mathbf{R}^J)\}$ for general $W \in \text{Gr}_m(\mathbf{R}^p)$, or (more importantly), a formula for $\max_{W \in \text{Gr}_m(\mathbf{R}^p)} \{\min_{J \in \mathcal{J}_m} \{d_{Gr}(W, \mathbf{R}^J)\}\}$.

Note that Proposition 5.6 asserts that statement 1 of Proposition 5.11 is true whenever $m \geq \frac{p}{2}$. To put into perspective the number $\frac{m}{8}\pi^2$ appearing in statement 2 of Proposition 5.11, and better understand the relevance of the comparison between m and $\frac{p}{2}$, note that the squared diameter of $\text{Gr}_m(\mathbf{R}^p)$ is $\min\{m, p-m\}\frac{\pi^2}{4}$. So for $m \leq \frac{p}{2}$, (5.15) is equivalent to

$$d_{Gr}(W, \mathbf{R}^J)^2 < \frac{1}{2} \text{diam}(\text{Gr}_m(\mathbf{R}^p))^2. \quad (5.16)$$

For $m > \frac{p}{2}$, the right-hand side of (5.15) is a greater fraction of $\text{diam}(\text{Gr}_m(\mathbf{R}^p))^2$, so it is “easier” for statement 2 of Proposition 5.11 to be true for $m > \frac{p}{2}$ than for $m < \frac{p}{2}$.

Remark 5.12. It is relatively easy to show that for any involution R , there exists σ for which RI_σ is not an involution. For $p = 2, 3$, we have $d_{SO}(I, R) \leq \pi$ for every $R \in SO(p)$, and $d_{SO}(I, R) = \pi$ for every involution R , so any non-involution is closer to the identity than is any involution. Hence for these values of p , Proposition 5.11 is easy to prove. However, for $p \geq 4$, given an involution R and a $\sigma \in \mathcal{I}_p^+$ for which RI_σ is not an involution, (5.10) shows that we cannot immediately deduce that $d_{SO}(I, RI_\sigma) < d_{SO}(I, R)$.

6. Proofs of the half-angle relation and results related to Grassmannians: Propositions 5.9, 5.10, and 5.11

The half-angle relation in Proposition 5.10 underlies our proofs of the most of the other results stated in Section 5.2 (all but Proposition 5.8). When the dimensions of the eigenspaces in Proposition 5.10 are equal, the half-angle relation leads to the elegant distance-relation (5.13). This equidimensional case is actually the only one we need for the application to Type-II nonuniqueness of MSSR curves. However, the half-angle relation (5.14) holds whether or not $\dim(E_{-1}(R_1)) = \dim(E_{-1}(R_2))$. Since this fact may be of interest outside

the scope of this paper, and is not much harder to prove without the equal-dimensions restriction, we have stated (and will prove) the more general relation.

Section 6.1 is devoted to establishing Proposition 5.10. In Section 6.2, we apply this proposition to establish Propositions 5.9 and 5.11.

6.1. The half-angle relation

We start with some notation.

Notation 6.1.

1. For $1 \leq i \leq p$ let \mathbf{e}_i denote the i^{th} standard basis vector of \mathbf{R}^p .
2. For $0 \leq m \leq p$, let $\mathcal{J}_{m,p}$ denote the collection of m -element subsets of $\{1, \dots, p\}$.
 - (a) For $0 \leq m \leq p$ and $J \in \mathcal{J}_{m,p}$, define $\boldsymbol{\sigma}^J = (\sigma_1, \dots, \sigma_p) \in \mathcal{I}_p$ by $\sigma_i = -1$ for $i \in J$ and $\sigma_i = 1$ for $i \notin J$. Similarly, for $\boldsymbol{\sigma} = (\sigma_1, \dots, \sigma_p) \in \mathcal{I}_p$, define $J^\sigma = \{i \in \{1, \dots, p\} : \sigma_i = -1\}$. (The maps $J \mapsto \boldsymbol{\sigma}^J$ and $\boldsymbol{\sigma} \mapsto J^\sigma$ are inverse to each other.)
 - (b) If $1 \leq m \leq p$ and $J = \{i_1, \dots, i_m\} \in \mathcal{J}_{m,p}$, with $i_1 < i_2 < \dots < i_m$, let \mathbf{E}_J denote the $p \times m$ matrix whose k^{th} column is \mathbf{e}_{i_k} , $1 \leq k \leq m$.
 - (c) For $0 \leq m \leq p$ and $J \in \mathcal{J}_{m,p}$, define $\mathbf{R}^J = \{(x^1, x^2, \dots, x^p) \in \mathbf{R}^p : x^i = 0 \text{ if } i \notin J\}$.

The collection $\{\mathbf{R}^J : J \in \mathcal{J}_{m,p}\}$ is the set of “coordinate m -planes” in \mathbf{R}^p .

3. For any $J \subset \{1, \dots, p\}$, let J' denote the complement of J in $\{1, \dots, p\}$.
4. For $m_1, m_2 \in \{1, 2, \dots, p\}$, $W \in \text{Gr}_{m_1}(\mathbf{R}^p)$, $Z \in \text{Gr}_{m_2}(\mathbf{R}^p)$, and $J \in \mathcal{J}_{m_1,p}$, writing $m = \min\{m_1, m_2\}$,
 - (a) let $\phi_1(W, Z), \dots, \phi_m(W, Z)$, denote the principal angles between the m_1 -plane W and the m_2 -plane Z (see [7, Section 12.4.3]), and
 - (b) let $\phi_{J,i}(W) = \phi_i(W, \mathbf{R}^J)$, $1 \leq i \leq m_1$.
5. For $1 \leq m \leq p$ define $d_{Gr} : \text{Gr}_m(\mathbf{R}^p) \times \text{Gr}_m(\mathbf{R}^p) \rightarrow \mathbf{R}$ by

$$d_{Gr}(W, Z) = \left(\sum_{i=1}^m \phi_i(W, Z)^2 \right)^{1/2}. \quad (6.1)$$

As noted earlier, d_{Gr} is the distance-function defined by the standard $SO(p)$ -invariant Riemannian metric on $\text{Gr}_m(\mathbf{R}^p)$ (up to a constant factor).

The following long but far-reaching technical lemma, giving several detailed relations between a general involution in $SO(p)$ and its product with a sign-change matrix, is our key tool for establishing the results stated in Section 5.2. It is best thought of as a series of lemmas, all with the same hypotheses, that have been rolled into one long lemma in order to avoid restating hypotheses and notational definitions. After proving the lemma, we build on it with two corollaries, completing the groundwork for the proofs (in later sections) of the Section 5.2 propositions.

Lemma 6.2. *Let $R \in SO(p)$ be an involution, let $\sigma \in \mathcal{I}_p^+$, assume $0 < m_\sigma := \text{level}(\sigma) < p$, and let $J = J^\sigma$ (see Notation 6.1). Viewing \mathbf{R}^p as $\mathbf{R}^{J'} \oplus \mathbf{R}^J$, below we write every $p \times p$ matrix in the block form $\begin{bmatrix} A_1 & A_2 \\ A_3 & A_4 \end{bmatrix}$, where A_1 is $(p - m_\sigma) \times (p - m_\sigma)$, A_2 is $(p - m_\sigma) \times m_\sigma$, A_3 is $m_\sigma \times (p - m_\sigma)$, and A_4 is $m_\sigma \times m_\sigma$. Then:*

(i) *In this block form,*

$$R = \begin{bmatrix} R_1 & R_2 \\ R_2^T & R_4 \end{bmatrix}, \quad (6.2)$$

where R_1 is a symmetric $(p - m_\sigma) \times (p - m_\sigma)$ matrix, R_4 is a symmetric $m_\sigma \times m_\sigma$ matrix, and R_2 is $(p - m_\sigma) \times m_\sigma$.

(ii) *In the same block form,*

$$(RI_\sigma)_{\text{sym}} = \frac{1}{2}(RI_\sigma + I_\sigma R) = \begin{bmatrix} R_1 & 0 \\ 0 & -R_4 \end{bmatrix}. \quad (6.3)$$

(iii) *All eigenvalues of R_1 and R_4 lie in the interval $[-1, 1]$.*

(iv) *For every $\lambda \in (-1, 1)$, if λ is an eigenvalue of R_1 (respectively, R_4), then $-\lambda$ is an eigenvalue of R_4 (resp. R_1) with the same multiplicity.*

(v) *Let l denote the number of eigenvalues of R_1 , counted with multiplicity, lying in the interval $(-1, 1)$. Then l is also the number of eigenvalues of R_4 , counted with multiplicity, lying in $(-1, 1)$, and $l \leq \min\{m_\sigma, p - m_\sigma\}$.*

(vi) *The inclusion map $\mathbf{R}^{J'} \rightarrow \mathbf{R}^p$ defined by $v \mapsto \begin{bmatrix} v \\ 0 \end{bmatrix}$ restricts to isomorphisms $E_{\pm 1}(R_1) \rightarrow E_{\pm 1}(R) \cap \mathbf{R}^{J'}$. Similarly the inclusion map $\mathbf{R}^J \rightarrow \mathbf{R}^p$ defined by $w \mapsto \begin{bmatrix} 0 \\ w \end{bmatrix}$ restricts to isomorphisms $E_{\pm 1}(R_4) \rightarrow E_{\pm 1}(R) \cap \mathbf{R}^J$.*

(vii) *Let $l_- = \dim(E_1(R) \cap \mathbf{R}^J)$, $l_+ = \dim(E_{-1}(R) \cap \mathbf{R}^{J'})$.³ Then $\dim(E_1(R_4)) = l_-$ and $\dim(E_{-1}(R_1)) = l_+$. (Thus $l_- + l_+$ is the multiplicity of -1 as an eigenvalue of $(RI_\sigma)_{\text{sym}}$ in (6.3), hence of RI_σ itself, and therefore yields a lower bound on $d_{SO}(RI_\sigma, I)$.) Furthermore,*

$$l_- \geq \text{level}(\sigma) - \text{level}(R) \quad \text{and} \quad l_+ \geq \text{level}(R) - \text{level}(\sigma), \quad (6.4)$$

³The \pm subscripts are chosen according to the eigenspaces of I_σ rather than R : $\mathbf{R}^J = E_{-1}(I_\sigma)$, $\mathbf{R}^{J'} = E_1(I_\sigma)$.

and

$$l_- - l_+ = \text{level}(\sigma) - \text{level}(R). \quad (6.5)$$

(viii) There exist an orthonormal R_1 -eigenbasis $\{v_i\}_{i=1}^{p-m_\sigma}$ of \mathbf{R}^{p-m_σ} (i.e. an orthonormal basis of \mathbf{R}^{p-m_σ} consisting of eigenvectors of R_1) and an R_4 -eigenbasis $\{w_i\}_{i=1}^{m_\sigma}$ of \mathbf{R}^{m_σ} . For any such bases $\{v_i\}$ of \mathbf{R}^{p-m_σ} , $\{w_i\}$ of \mathbf{R}^{m_σ} , let $\{\lambda'_i\}, \{\lambda_i\}$ be the corresponding eigenvalues (i.e. $R_1 v_i = \lambda'_i v_i$ and $R_4 w_i = \lambda_i w_i$), and define

$$\mathbf{v}_i = \begin{cases} \begin{bmatrix} \frac{1}{1+\lambda'_i} R_2^T v_i \\ v_i \end{bmatrix}, & 1 < i \leq p - m_\sigma, \lambda'_i \neq -1, \\ \begin{bmatrix} v_i \\ 0 \end{bmatrix}, & 1 \leq i \leq p - m_\sigma, \lambda'_i = -1, \end{cases} \quad (6.6)$$

$$\mathbf{w}_i = \begin{cases} \begin{bmatrix} \frac{-1}{1-\lambda_i} R_2 w_i \\ w_i \end{bmatrix}, & 1 \leq i \leq m_\sigma, \lambda_i \neq 1, \\ \begin{bmatrix} 0 \\ w_i \end{bmatrix}, & 1 \leq i \leq m_\sigma, \lambda_i = 1. \end{cases} \quad (6.7)$$

Then

$$\left\{ \sqrt{\frac{1+\lambda'_i}{2}} \mathbf{v}_i : 1 \leq i \leq p - m_\sigma, \lambda'_i \neq -1 \right\} \cup \{ \mathbf{w}_i : 1 \leq i \leq m_\sigma, \lambda_i = 1 \} \quad (6.8)$$

(ordered arbitrarily) is an orthonormal basis of $E_1(R)$, and the set

$$\left\{ \sqrt{\frac{1-\lambda_i}{2}} \mathbf{w}_i : 1 \leq i \leq m_\sigma, \lambda_i \neq 1 \right\} \cup \{ \mathbf{v}_i : 1 \leq i \leq p - m_\sigma, \lambda'_i = -1 \} \quad (6.9)$$

(ordered arbitrarily) is an orthonormal basis of $E_{-1}(R)$. Note that the cardinality of the second set in (6.8) (respectively (6.9)) is l_- (resp. l_+).

Proof: To simplify notation in this proof, we let $m = m_\sigma$.

Since $R \in SO(p)$ is an involution, $R = R^{-1} = R^T$. Hence R is symmetric, implying assertion (i), and \mathbf{R}^p is the orthogonal direct sum of $E_1(R)$ and $E_{-1}(R)$ (since the only possible eigenvalues of an involution are ± 1).

For (ii), observe that in the block-form decomposition we are using,

$$I_\sigma = \begin{bmatrix} I_{(p-m) \times (p-m)} & 0 \\ 0 & -I_{m \times m} \end{bmatrix}.$$

A simple calculation then yields (6.3).

Next, because $R^2 = I$, we have the following relations:

$$R_1^2 + R_2 R_2^T = I_{(p-m) \times (p-m)}, \quad (6.10)$$

$$R_1 R_2 + R_2 R_4 = 0_{(p-m) \times m}, \quad (6.11)$$

$$R_2^T R_1 + R_4 R_2^T = 0_{m \times (p-m)}, \quad (6.12)$$

$$R_4^2 + R_2^T R_2 = I_{m \times m}. \quad (6.13)$$

From (6.10) and (6.13), for any $v \in \mathbf{R}^{p-m}, w \in \mathbf{R}^m$, we have

$$\|R_1 v\|^2 + \|R_2^T v\|^2 = \|v\|^2, \quad (6.14)$$

$$\|R_4 w\|^2 + \|R_2 w\|^2 = \|w\|^2. \quad (6.15)$$

It follows from (6.14)–(6.15) that if λ is an eigenvalue of R_1 or R_4 , then $|\lambda| \leq 1$, yielding (iii).

To obtain (iv), consider the operators $L : \mathbf{R}^m \rightarrow \mathbf{R}^{p-m}$ and $L^* : \mathbf{R}^{p-m} \rightarrow \mathbf{R}^m$ defined by $L(w) = R_2 w$ and $L^*(v) = R_2^T v$. Suppose that R_1 has an eigenvalue λ with $|\lambda| < 1$, and let $0 \neq v \in E_\lambda(R_1)$. Let $w = R_2^T v$; note that (6.14) implies $w \neq 0$. Using (6.12),

$$R_4 w = R_4 R_2^T v = -R_2^T R_1 v = -R_2^T \lambda v = -\lambda w.$$

Hence L^* maps $E_\lambda(R_1)$ injectively to $E_{-\lambda}(R_4)$. Similarly, if R_4 has an eigenvalue $-\lambda$ with $|\lambda| < 1$, and L^* maps $E_{-\lambda}(R_4)$ injectively to $E_\lambda(R_1)$.

It follows that, for any $\lambda \in \mathbf{R}$ with $|\lambda| < 1$, λ is an eigenvalue of R_1 if and only if $-\lambda$ is an eigenvalue of R_4 , and that the maps

$$L^*|_{E_\lambda(R_1)} : E_\lambda(R_1) \rightarrow E_{-\lambda}(R_4) = E_\lambda(-R_4), \quad (6.16)$$

$$L|_{E_\lambda(-R_4)} : E_\lambda(-R_4) = E_{-\lambda}(R_4) \rightarrow E_\lambda(R_1) \quad (6.17)$$

are isomorphisms. This establishes (iv). Statement (v) is an immediate corollary of (iv).

For (vi), let $\iota : \mathbf{R}^{J'} \rightarrow \mathbf{R}^p$ be the first inclusion map in the lemma. Note that $R \begin{bmatrix} v \\ 0 \end{bmatrix} = \begin{bmatrix} R_1 v \\ R_2^T v \end{bmatrix}$. If $v \in E_\lambda(R_1)$ with $\lambda = \pm 1$, equation (6.14) implies that $R_2^T v = 0$, hence that $R\iota(v) = \lambda\iota(v)$. Conversely, if $R\iota(v) = \lambda\iota(v)$, then $R_1 v = \lambda v$ (and $R_2^T v = 0$). Hence ι carries $E_\lambda(R_1)$ isomorphically to $E_\lambda(R) \cap \mathbf{R}^{J'}$. The argument for the inclusion map $\mathbf{R}^J \rightarrow \mathbf{R}^p$ is essentially identical. This establishes (vi).

Part (vi) implies that $\dim(E_1(R_4)) = \dim(E_1(R) \cap \mathbf{R}^J) = l_-$ and that $\dim(E_{-1}(R_1)) = \dim(E_{-1}(R) \cap \mathbf{R}^{J'}) = l_+$, the first assertion in (vii). To obtain (6.4)–(6.5), note that for any subspaces V, W of \mathbf{R}^p , we have

$$\dim(V^\perp \cap W) - \dim(V \cap W^\perp) = \dim(W) - \dim(V). \quad (6.18)$$

(The proof of (6.18) is straightforward linear algebra.) Applying this to the case $V = E_{-1}(R)$, $V^\perp = E_1(R)$, $W = E_{-1}(I_\sigma) = \mathbf{R}^J$, $W^\perp = E_1(I_\sigma) = \mathbf{R}^{J'}$, we have $l_- = \dim(V^\perp \cap W)$ and $l_+ = \dim(V \cap W^\perp)$, so (6.5) follows from (6.18). The inequalities in (6.4) follow directly from (6.5).

(viii) Since R_1 (respectively R_4) is symmetric, an orthonormal R_1 -eigenbasis $\{v_i\}$ of \mathbf{R}^{p-m} (resp., orthonormal R_4 -eigenbasis $\{w_i\}$ of \mathbf{R}^m) exists. Select such eigenbases, and let $\{\lambda_i\}$, $\{\lambda'_i\}$ be eigenvalues as defined in the Lemma. Note that the second set in (6.8) is a basis of $E_1(R_4)$, which by (vi) is isomorphic to $E_1(R) \cap \mathbf{R}^J$. Hence the cardinality of this set is $\dim(E_1(R) \cap \mathbf{R}^J)$, i.e. l_- . Similarly, the second set in (6.9) is a basis of $E_{-1}(R_1)$ and has cardinality l_+ .

Without loss of generality, we may assume that the eigenvectors v_i with eigenvalue -1 , if any, are the last l_+ , and that the eigenvectors w_i with eigenvalue 1 , if any, are the last l_- . Using (6.13), for $1 \leq i \leq m$ we have $R_2^T R_2 w_i = (1 - \lambda_i^2)w_i$, while using (6.11) we find $R_1 R_2 w_i = -\lambda_i R_2 w_i$. Then, using (6.2), a simple calculation shows that $R w_i = -w_i$. Hence $w_i \in E_{-1}(R)$ for $1 \leq i \leq m - l_-$, while from part (vi), $v_i \in E_{-1}(R)$ for $p - m - l_+ < i \leq p - m$.

Let $\langle \cdot, \cdot \rangle$ denote the standard inner product on \mathbf{R}^n for any n . As seen in the proof of part (vi), $v \in E_{-1}(R_1)$ implies $R_2^T v = 0$. Hence for $p - m - l_+ < i \leq p - m$ and $1 \leq j \leq m - l_-$, $\langle v_i, w_j \rangle \propto \langle v_i, R_2 w_j \rangle = \langle R_2^T v_i, w_j \rangle = 0$, while for $p - m - l_+ < i, j \leq p - m$ we have $\langle v_i, v_j \rangle = \langle v_i, v_j \rangle = \delta_{ij}$. Finally, for $i, j \leq m - l_-$, using the fact that $\langle R_2 w_i, R_2 w_j \rangle = \langle w_i, R_2^T R_2 w_j \rangle = \langle w_i, (1 - \lambda_i^2)w_j \rangle$, a simple computation yields $\langle w_i, w_j \rangle = \frac{2}{1 - \lambda_i^2} \delta_{ij}$. Thus $\{\sqrt{\frac{1 - \lambda_i^2}{2}} w_i : 1 \leq i \leq m - l_-\} \cup \{v_i : p - m - l_+ < i \leq p - m\}$ is an orthonormal subset of $E_{-1}(R)$. Using (6.5), the cardinality of this subset is $m - l_- + l_+ = \text{level}(\sigma) - (\text{level}(\sigma) - \text{level}(R)) = \text{level}(R) = \dim(E_{-1}(R))$. Hence (6.9) is an orthonormal basis of $E_{-1}(R)$.

The proof that (6.8) is an orthonormal basis of $E_1(R)$ is similar. \blacksquare

Corollary 6.3. *Hypotheses and notation as in Lemma 6.2. Let $l_+ = \dim(E_{-1}(R_1))$ and $l_- = \dim(E_1(R_4))$ (as in Lemma 6.2(vii)). In addition let $\{\theta_i \in [0, \pi]\}_{i=1}^{\lceil p/2 \rceil}$ be angles for which $R(\theta_1, \dots, \theta_{\lceil p/2 \rceil})$ is a normal form of RI_σ , and let $\{\tilde{\theta}_i\}_{i=1}^p$ be as defined in (5.11). Let $J_* = \{j \in J : 0 < \tilde{\theta}_j < \pi\}$. Then $|J_*| \leq \min\{m_\sigma, p - m_\sigma\}$, and*

$$d_{SO}(RI_\sigma, I)^2 = \frac{1}{2}(l_+ + l_-)\pi^2 + \sum_{j \in J_*} \tilde{\theta}_j^2. \quad (6.19)$$

If $\text{level}(\sigma) = \text{level}(R)$, then

$$d_{SO}(RI_\sigma, I)^2 = l_- \pi^2 + \sum_{j \in J_*} \tilde{\theta}_j^2 = \sum_{j \in J} \tilde{\theta}_j^2. \quad (6.20)$$

Proof: Let $\beta' : J' \rightarrow \{1, \dots, p - m_\sigma\}$, $\beta : J \rightarrow \{1, \dots, m_\sigma\}$, be order-preserving bijections. By (5.8), the eigenvalues of $(RI_\sigma)_{\text{sym}}$, counted with multiplicity, are

$\{\cos \tilde{\theta}_i\}_{i=1}^p$. But from Lemma 6.2(ii), we can read off the eigenvalues of $(RI_{\sigma})_{\text{sym}}$ from (6.3); they are $\lambda'_1, \dots, \lambda'_{p-m_{\sigma}}, -\lambda_1, \dots, -\lambda_{m_{\sigma}}$ (ordered arbitrarily). Thus, reordering the λ'_j and the λ_j appropriately, for $1 \leq j \leq p$ we have

$$\cos \tilde{\theta}_j = \begin{cases} \lambda'_{\beta'(j)} & \text{if } j \in J', \\ -\lambda_{\beta(j)} & \text{if } j \in J. \end{cases} \quad (6.21)$$

Define $J_* = \{j \in J : \lambda_{\beta(j)} \neq \pm 1\}$. Observe that J_* can also be characterized as $\{j \in J : \lambda_{\beta(j)} \neq \pm 1\}$. Similarly, define $J'_* = \{j \in J' : \lambda'_{\beta'(j)} \neq \pm 1\} = \{j \in J' : 0 < \tilde{\theta}_j < \pi\}$. By part (v) of Lemma 6.2, $|J'_*| = |J_*| = l \leq \min\{m_{\sigma}, p - m_{\sigma}\}$, and by part (iv) of the Lemma there is a bijection $b : J_* \rightarrow J'_*$ such that $-\lambda_j = \lambda'_{b(j)}$ for all $j \in J_*$. Hence

$$\tilde{\theta}_j = \begin{cases} \cos^{-1} \lambda'_{\beta'(j)} & \text{if } j \in J'_*, \\ \cos^{-1} \lambda'_{b(\beta(j))} & \text{if } j \in J_*, \\ 0 \text{ or } \pi & \text{otherwise.} \end{cases} \quad (6.22)$$

In particular,

$$\sum_{j \in J_*} \tilde{\theta}_j^2 = \sum_{j \in J'_*} \tilde{\theta}_j^2. \quad (6.23)$$

Next, note that

$$\sum_{j \in J' \setminus J'_*} \tilde{\theta}_j^2 = \sum_{\{j \in J' : \tilde{\theta}_j = \pi\}} \tilde{\theta}_j^2 = \#\{j \in J' : \lambda'_{\beta'(j)} = -1\} \pi^2 = \dim(E_{-1}(R_1)) \pi^2 = l_+ \pi^2, \quad (6.24)$$

and similarly $\sum_{j \in J \setminus J_*} \tilde{\theta}_j^2 = l_- \pi^2$. From (5.12) we therefore have

$$\begin{aligned} d_{SO}(RI_{\sigma}, I)^2 &= \frac{1}{2} \left\{ \sum_{j \in J' \setminus J'_*} \tilde{\theta}_j^2 + \sum_{j \in J \setminus J_*} \tilde{\theta}_j^2 + \sum_{j \in J'_*} \tilde{\theta}_j^2 + \sum_{j \in J_*} \tilde{\theta}_j^2 \right\} \\ &= \frac{1}{2} \left\{ l_+ \pi^2 + l_- \pi^2 + 2 \sum_{j \in J_*} \tilde{\theta}_j^2 \right\}, \end{aligned}$$

establishing (6.19).

If $\text{level}(\sigma) = \text{level}(R)$, then equation (6.5) implies that $l_+ = l_-$, so (6.19) implies the first equality in (6.20). For the second equality, observe that $j \in J \setminus J_*$ if and only if $\tilde{\theta}_j$ is 0 or π . The number of j 's in J for which $\tilde{\theta}_j = \pi$ is exactly l_- , while the j 's in J for which $\tilde{\theta}_j = 0$ have no effect on $\sum_{j \in J} \tilde{\theta}_j^2$. Hence the second equality in (6.20) holds. \blacksquare

Corollary 6.4. *Hypotheses and notation as in Lemma 6.2, except that we additionally write $m_R := \text{level}(R)$ and $m = \min\{m_\sigma, m_R\}$. Let $\{\theta_i \in [0, \pi]\}_{i=1}^{\lceil p/2 \rceil}$ be angles for which $R(\theta_1, \dots, \theta_{\lceil p/2 \rceil})$ is a normal form of RI_σ , let $\{\tilde{\theta}_i\}_{i=1}^p$ be as defined in (5.11), let the elements of J be $i_1 < i_2 < \dots < i_{m_\sigma}$, and let $\phi_{J,j} = \phi_{J,j}(E_{-1}(R))$, $1 \leq j \leq m$. Then:*

(i) *Up to ordering,*

$$\phi_{J,j} = \frac{\tilde{\theta}_{i_j}}{2}, \quad 1 \leq j \leq m. \quad (6.25)$$

(ii) *If $m_\sigma = m_R$ then*

$$d_{SO}(RI_\sigma, I) = 2d_{Gr}(E_{-1}(R), \mathbf{R}^J). \quad (6.26)$$

Proof: (i). Let \widetilde{W} be the $p \times m_R$ matrix formed by the columns of the basis (6.9) of $E_{-1}(R)$, with the elements of the first set in (6.9) comprising the first $m_\sigma - l_-$ columns, and the elements of the second set comprising the last l_+ columns. (Here l_\pm are defined as in Lemma 6.2(vii).) Without loss of generality we order the R_4 -eigenvectors w_i such that the first $m_\sigma - l_-$ are the ones for which $\lambda_i \neq 1$.

Since the columns of \widetilde{W} form an orthonormal basis of $E_{-1}(R)$, the numbers $\{\cos \phi_{J,i}\}_{i=1}^m$ are the singular values of the $m_R \times m_\sigma$ matrix $\widetilde{W}^T E_J$. (This is true whether $m_R \leq m_\sigma$ or $m_R > m_\sigma$.) But, relative to the block-decomposition of matrices used in Lemma 6.2, the upper $(p - m_\sigma) \times m_\sigma$ block of E_J is 0, and the lower $m_\sigma \times m_\sigma$ block is $I_{m_\sigma \times m_\sigma}$. Hence, writing \widetilde{W}_* for the $m_\sigma \times (m_\sigma - l_-)$ matrix formed by the last m_σ rows of the first $m_\sigma - l_-$ columns of \widetilde{W} , and noting that $m_\sigma - l_- = m_R - l_+$ (by (6.5)), we have $\widetilde{W}^T E_J = \begin{bmatrix} \widetilde{W}_*^T \\ 0_{l_+ \times m_\sigma} \end{bmatrix}$, where the i^{th} row of the $(m_\sigma - l_-) \times (p - m_\sigma)$ matrix \widetilde{W}_*^T is a multiple of w_i^T . Hence for $i, j \leq m_\sigma - l_- = m_R - l_+$,

$$\left((\widetilde{W}^T E_J)(\widetilde{W}^T E_J)^T \right)_{ij} = \sqrt{\frac{1 - \lambda_i}{2}} \sqrt{\frac{1 - \lambda_j}{2}} \langle w_i, w_j \rangle = \frac{1 - \lambda_j}{2} \delta_{ij}, \quad (6.27)$$

and all other entries of the $m_R \times m_R$ matrix $\widetilde{W}^T E_J (\widetilde{W}^T E_J)^T$ are 0. But for $m_\sigma - l_- < i \leq m_\sigma$, we have $\lambda_i = 1$, so $\left((\widetilde{W}^T E_J)(\widetilde{W}^T E_J)^T \right)_{ij} = \frac{1 - \lambda_j}{2} \delta_{ij}$ for all $i, j \leq m = \min\{m_R, m_\sigma\}$. Thus the upper left-hand $m \times m$ block of $(\widetilde{W}^T E_J)(\widetilde{W}^T E_J)^T$ (the entire $m_R \times m_R$ matrix if $m_R \leq m_\sigma$) is $\text{diag}(\frac{1 - \lambda_1}{2}, \dots, \frac{1 - \lambda_m}{2})$, so the numbers $\sqrt{\frac{1 - \lambda_j}{2}}$, $1 \leq j \leq m$, are the singular values of $\widetilde{W}^T E_J$. Thus, up to ordering, the principal angles $\{\phi_{J,i}\}$ are given by

$$\cos \phi_{J,j} = \sqrt{\frac{1 - \lambda_j}{2}}, \quad 1 \leq j \leq m. \quad (6.28)$$

The bijection $\beta : J \rightarrow \{1, \dots, m_{\sigma}\}$ used in the proof of Corollary 6.3 is simply the inverse of the map $j \mapsto i_j$. Thus from (6.21), we have

$$-\lambda_j = \cos \tilde{\theta}_{i_j}, \quad 1 \leq j \leq m_{\sigma}. \quad (6.29)$$

Combining (6.28) with (6.29),

$$\cos \phi_{J,j} = \sqrt{\frac{1 + \cos \tilde{\theta}_{i_j}}{2}} = \cos \frac{\tilde{\theta}_{i_j}}{2}. \quad (6.30)$$

But $\tilde{\theta}_{i_j} \in [0, \pi]$, so both $\phi_{J,j}$ and $\frac{\tilde{\theta}_{i_j}}{2}$ lie in $[0, \frac{\pi}{2}]$. Hence (6.30) implies that $\phi_{J,j} = \tilde{\theta}_{i_j}/2$, $1 \leq j \leq m$.

(ii) Assume $m_{\sigma} = m_R$; then both equal m . Corollary 6.3 then implies that

$$d_{SO}(RI_{\sigma}, I)^2 = \sum_{i=1}^m \tilde{\theta}_{i_j}^2. \quad (6.31)$$

But from part (i) we have $\tilde{\theta}_{i_j} = 2\phi_{J,j}$ for $1 \leq j \leq m$, so, using (6.20), $d_{SO}(RI_{\sigma}, I)^2 = 4d_{Gr}(E_{-1}(R), \mathbf{R}^J)^2$, implying (6.26). ■

We are now ready to establish the general half-angle relation:

Proof of Proposition 5.10. Let $U \in O(p)$ and let $T_U : \mathbf{R}^p \rightarrow \mathbf{R}^p$ be the corresponding orthogonal transformation. For any even $m' > 0$ and any $R \in \text{Inv}_{m'}(p)$, we have

$$E_{-1}(URU^{-1}) = T_U(E_{-1}(R)). \quad (6.32)$$

Now let $T : \mathbf{R}^p \rightarrow \mathbf{R}^p$ be an orthogonal transformation carrying $E_{-1}(R_2)$ to a coordinate plane \mathbf{R}^J , and let $U \in O(p)$ be the matrix for which $T = T_U$. Then $UR_2U^{-1} = I_{\sigma}$, where $\sigma = \sigma^J$. For $i = 1, 2$ let $R'_i = UR_iU^{-1}$. Since T is an orthogonal transformation, the (multi-)set of principal angles between $E_{-1}(R'_1) = T(E_{-1}(R_1))$ and $E_{-1}(R'_2) = T(E_{-1}(R_2))$ is identical to the (multi-)set of principal angles between $E_{-1}(R_1)$ and $E_{-1}(R_2)$. But $R'_1 I_{\sigma} = R'_1 R'_2 = UR_1 R_2 U^{-1}$, so $R(\theta_1, \dots, \theta_{\lceil p/2 \rceil})$ is a normal form of $R'_1 I_{\sigma}$ as well as of $R_1 R_2$. The result now follows from Corollary 6.4(i) and equation (6.22) (the latter being needed only for the final statement of the result). ■

6.2. The proofs of Propositions 5.9 and 5.11

Proof of Proposition 5.9. Since $d_{SO}(RI_{\sigma}, I) = d_{SO}(R, I_{\sigma}^{-1}) = d_{SO}(R, I_{\sigma})$, conclusion (ii) of Corollary 6.4 can be written equivalently as:

$$d_{SO}(\Phi(W), \Phi(\mathbf{R}^J)) = 2d_{Gr}(W, \mathbf{R}^J). \quad (6.33)$$

Fix any $J \in \mathcal{J}_{m,p}$. Letting “ \cdot ” denote the natural left-action of $SO(p)$ on $\text{Gr}_m(\mathbf{R}^p)$ ($U \cdot W = T_U(W)$), observe that, in the notation of the proof of

Proposition 5.10), for all $U \in SO(p)$ and $W \in \text{Gr}_m(\mathbf{R}^p)$ we have $\Phi(U \cdot W) = U\Phi(W)U^{-1}$ (simply another way of writing (6.32).) Clearly d_{Gr} is invariant under this action, and d_{SO} is both left- and right-invariant, so (6.33) implies that

$$d_{SO}(\Phi(U \cdot W), \Phi(U \cdot \mathbf{R}^J)) = 2d_{Gr}(U \cdot W, U \cdot \mathbf{R}^J).$$

Now let $W, V \in \text{Gr}_m(\mathbf{R}^p)$. Since the action of $SO(p)$ on $\text{Gr}_m(\mathbf{R}^p)$ is transitive, there exists $U \in SO(p)$ such that $U \cdot \mathbf{R}^J = V$. Using any such U , we then have

$$\begin{aligned} d_{SO}(\Phi(W), \Phi(V)) &= d_{SO}(\Phi(U \cdot U^{-1} \cdot W), \Phi(U \cdot \mathbf{R}^J)) \\ &= 2d_{Gr}(U \cdot U^{-1} \cdot W, U \cdot \mathbf{R}^J) \\ &= 2d_{Gr}(W, V). \end{aligned}$$

■

Remark 6.5. Of course, Proposition 5.9 can be deduced from computations with the principal fibration

$$\pi : SO(p) \rightarrow SO(p)/S(O(m) \times O(p-m)) \cong \text{Gr}_m(\mathbf{R}^p);$$

the standard Riemannian metric on $\text{Gr}_m(\mathbf{R}^p)$ (for which d_{Gr} is the geodesic-distance function) is defined so as to make π a Riemannian submersion up to a normalization constant. Our proof of Proposition 5.9 is independent of this Riemannian proof in the sense that it establishes equality between the left-hand side of (6.33) and the right-hand side *as defined by equation* (6.1). Without the *a priori* knowledge that d_{Gr} is a geodesic-distance function, it is not obvious that d_{Gr} satisfies the triangle inequality, hence whether d_{Gr} is a metric. Thus Proposition 5.9 actually provides an independent proof that d_{Gr} is a metric on $\text{Gr}_m(\mathbf{R}^p)$. The only use of Riemannian geometry in this proof is through the knowledge that d_{SO} is, in fact, a metric (because it is a geodesic-distance function).

Proof of Proposition 5.11. Let “Statement 1” and “Statement 2” be the statements listed as 1 and 2 in the Proposition. As noted in the proof of Proposition 5.9, $d_{SO}(RI_\sigma, I) = d_{SO}(R, I_\sigma)$, so the inequality $d_{SO}(RI_\sigma, I) < d_{SO}(R, I)$ can be rewritten as

$$d_{SO}(R, I_\sigma)^2 < \frac{m\pi^2}{2}.$$

Assume first that Statement 1 is true. Let $W \in \text{Gr}_m(\mathbf{R}^p)$. Then $\Phi_{m,p}(W)$ is an involution of level m , so there exists $\sigma \in \mathcal{I}_p^+$ of level m such that $d_{SO}(\Phi_{m,p}(W), I_\sigma)^2 < \frac{m\pi^2}{2}$. Select such a σ and let $J = J^\sigma$. Then $I_\sigma = \Phi_{m,p}(\mathbf{R}^J)$, so

$$\begin{aligned}
d_{Gr}(W, \mathbf{R}^J)^2 &= \frac{1}{4} d_{SO}(\Phi_{m,p}(W), \Phi_{m,p}(\mathbf{R}^J))^2 = \frac{1}{4} d_{SO}(\Phi_{m,p}(W), I_\sigma)^2 \\
&< \frac{m\pi^2}{8} .
\end{aligned}$$

Hence Statement 2 is true.

Conversely, assume that Statement 2 is true. Let $R \in \text{Inv}_m(p)$. Then there exists $J \in \mathcal{J}_{m,p}$ such that $d_{Gr}(\Phi_{m,p}^{-1}(R), \mathbf{R}^J)^2 < \frac{m\pi^2}{8}$. Select such a J and let $\sigma = \sigma^J$. Then $I_\sigma = \Phi_{m,p}(\mathbf{R}^J)$, so

$$d_{SO}(R, I_\sigma)^2 = d_{SO}(R, \Phi_{m,p}(\mathbf{R}^J))^2 = 4d_{Gr}(\Phi_{m,p}^{-1}(R), \mathbf{R}^J)^2 < \frac{m\pi^2}{2} .$$

Hence Statement 1 is true. ■

7. Proofs of sign-change reducibility results, part I: Propositions 5.6 and 5.8

We are now ready to attack the question of sign-change reducibility: given $R \in \text{Inv}(p)$, can we find $\sigma \in \mathcal{I}_p^+$ such that $d_{SO}(RI_\sigma, I) < d_{SO}(R, I)$? Equations (5.9) and (6.19) tell us that this inequality is satisfied if and only if

$$(l_+ + l_-)\pi^2 + 2 \sum_{j \in J_*} \tilde{\theta}_j^2 < \text{level}(R)\pi^2, \quad (7.1)$$

where $l_\pm = l_\pm(R, \sigma)$ are as in Lemma 6.2(vii). Since π is the largest possible value for a normal-form angle in (5.2), it is reasonable to try to look for a σ such that l_+ and l_- are as small as possible. However, to achieve (7.1), we have to make sure that we do not make $\sum_{j \in J_*} \tilde{\theta}_j^2$ too large while we are making l_\pm small. We next prove a lemma that, via its subsequent corollary, will help us show that for $\text{level}(R) = m \geq \frac{p}{2}$, we can choose $J \in \mathcal{J}_{m,p}$ to make $d_{SO}(RI_\sigma, I)$ as small as is needed to prove Proposition 5.6.

Lemma 7.1. *For $1 \leq m \leq p$,*

$$\sum_{J \in \mathcal{J}_{m,p}} \mathbf{E}_J \mathbf{E}_J^T = \binom{p-1}{m-1} I_{p \times p}. \quad (7.2)$$

Proof: For $J = (i_1, \dots, i_m) \in \mathcal{J}_{m,p}$, we have

$$\mathbf{E}_J \mathbf{E}_J^T = \sum_{i \in J} \mathbf{e}_i \mathbf{e}_i^T. \quad (7.3)$$

Hence when $m = 1$ and when $m = p$, the left-hand side of (7.2) reduces to $I_{p \times p}$, which is also true of the right-hand side.

We proceed by induction on p . For each $p \geq 1$, consider the statement

$$S(p) : \text{Equation (7.2) is true for all } m \text{ satisfying } 1 \leq m \leq p. \quad (7.4)$$

We have already established that (7.2) holds for $m = 1 = p$, hence that statement $S(1)$ is true. Now suppose that $S(p)$ is true for some given p . To consider $S(p+1)$, let $\{\mathbf{e}_i\}_{i=1}^p, \{\mathbf{e}'_i\}_{i=1}^{p+1}$ denote the standard bases of $\mathbf{R}^p, \mathbf{R}^{p+1}$ respectively. For $K = \{i_1, \dots, i_m\} \in \mathcal{J}_{p+1, m}$ with $i_1 < i_2 < \dots < i_m$ we write E'_K for the $(p+1) \times m$ matrix whose j^{th} column is \mathbf{e}'_{i_j} , $1 \leq j \leq m$. Note that

$$\begin{aligned} \mathbf{e}'_i &= \begin{bmatrix} \mathbf{e}_i \\ 0 \end{bmatrix} \quad \text{for } 1 \leq i \leq p, \\ E'_J(E'_J)^T &= \begin{bmatrix} E_J(E_J)^T & 0_{p \times 1} \\ 0_{1 \times p} & 0 \end{bmatrix} \quad \text{for } J \in \mathcal{J}_{m, p}, \\ \text{and} \quad \mathbf{e}'_{p+1}(\mathbf{e}'_{p+1})^T &= \begin{bmatrix} 0_{p \times p} & 0_{p \times 1} \\ 0_{1 \times p} & 1 \end{bmatrix}. \end{aligned}$$

Hence for $1 \leq m \leq p$,

$$\begin{aligned} & \sum_{K \in \mathcal{J}_{m, p+1}} E'_K(E'_K)^T \\ &= \sum_{\{K \in \mathcal{J}_{m, p+1} : p+1 \in K\}} E'_K(E'_K)^T + \sum_{\{K \in \mathcal{J}_{m, p+1} : p+1 \notin K\}} E'_K(E'_K)^T \\ &= \sum_{\{K \in \mathcal{J}_{m, p+1} : K = J \cup \{p+1\} \text{ for some } J \in \mathcal{J}_{m-1, p}\}} E'_K(E'_K)^T + \sum_{K \in \mathcal{J}_{m, p}} E'_K(E'_K)^T \\ &= \sum_{J \in \mathcal{J}_{m-1, p}} (E'_J(E'_J)^T + \mathbf{e}'_{p+1}(\mathbf{e}'_{p+1})^T) + \sum_{J \in \mathcal{J}_{m, p}} \begin{bmatrix} E_J(E_J)^T & 0_{p \times 1} \\ 0_{1 \times p} & 0 \end{bmatrix} \\ &= \sum_{J \in \mathcal{J}_{m-1, p}} \left(\begin{bmatrix} E_J E_J^T & 0_{p \times 1} \\ 0_{1 \times p} & 0 \end{bmatrix} \right) + |\mathcal{J}_{m-1, p}| \mathbf{e}'_{p+1}(\mathbf{e}'_{p+1})^T \\ &\quad + \sum_{J \in \mathcal{J}_{m, p}} \begin{bmatrix} E_J(E_J)^T & 0_{p \times 1} \\ 0_{1 \times p} & 0 \end{bmatrix} \\ &= \begin{bmatrix} \sum_{J \in \mathcal{J}_{m-1, p}} E_J E_J^T + \sum_{J \in \mathcal{J}_{m, p}} E_J E_J^T & 0_{p \times 1} \\ 0_{1 \times p} & 0 \end{bmatrix} + |\mathcal{J}_{m-1, p}| \mathbf{e}'_{p+1}(\mathbf{e}'_{p+1})^T \\ &= \begin{bmatrix} \left\{ \binom{p-1}{m-2} + \binom{p-1}{m-1} \right\} I_{p \times p} & 0_{p \times 1} \\ 0_{1 \times p} & 0 \end{bmatrix} + \binom{p}{m-1} \mathbf{e}'_{p+1}(\mathbf{e}'_{p+1})^T \\ &= \binom{p}{m-1} I_{(p+1) \times (p+1)} \end{aligned}$$

Hence (7.1) holds with p replaced by $p+1$, as long as $1 \leq m \leq p$. But we have already established that (7.1) holds whenever $m = p$; hence if p is replaced

by $p + 1$, the equality holds for $m = p + 1$. Thus (7.1) holds for all m with $1 \leq m \leq p + 1$; i.e. statement $S(p + 1)$ is true. By induction, $S(p)$ is true for all p , which is exactly what the Lemma asserts. ■

Corollary 7.2. *Let $m \in \{1, 2, \dots, p\}$ and let $W \in \text{Gr}_m(\mathbf{R}^p)$. There exists $J \in \mathcal{J}_{m,p}$ such that*

$$\sum_{i=1}^m \sin^2 \phi_{J,i} \leq m(1 - \frac{m}{p}). \quad (7.5)$$

Furthermore, the inequality in (7.5) is strict for some $J \in \mathcal{J}_{m,p}$ unless equality holds in (7.5) for all $J \in \mathcal{J}_{m,p}$.

Proof: Let \widetilde{W} be any $p \times m$ matrix whose columns are an orthonormal basis of W . Using Lemma 7.1,

$$\begin{aligned} \sum_{J \in \mathcal{J}_{m,p}} \text{tr}(\widetilde{W}^T \mathbf{E}_J \mathbf{E}_J^T \widetilde{W}) &= \text{tr} \left(\widetilde{W}^T \left(\sum_{J \in \mathcal{J}_{m,p}} \mathbf{E}_J \mathbf{E}_J^T \right) \widetilde{W} \right) \\ &= \text{tr} \left(\widetilde{W}^T \binom{p-1}{m-1} I_{p \times p} \widetilde{W} \right) \\ &= m \binom{p-1}{m-1} \end{aligned}$$

since $\widetilde{W}^T \widetilde{W} = I_{m \times m}$.

Since $|\mathcal{J}_{m,p}| = \binom{p}{m}$, the average of $\text{tr}(\widetilde{W}^T \mathbf{E}_J \mathbf{E}_J^T \widetilde{W})$ over all $J \in \mathcal{J}_{m,p}$ is $m \binom{p-1}{m-1} / \binom{p}{m} = m^2/p$. Hence $\text{tr}(\widetilde{W}^T \mathbf{E}_J \mathbf{E}_J^T \widetilde{W}) \geq m^2/p$ for at least one $J \in \mathcal{J}_{m,p}$, and the inequality is strict for some J unless it is an equality for all J . But for any $Z \in \text{Gr}_m(\mathbf{R}^p)$, the principal angles ϕ_1, \dots, ϕ_m between W and Z are the numbers in $[0, \frac{\pi}{2}]$ for which $\cos \phi_1, \dots, \cos \phi_m$ are the singular values of the $m \times m$ matrix $\widetilde{W}^T \widetilde{Z}$, where \widetilde{Z} is any $p \times m$ matrix whose columns are an orthonormal basis of Z . Since for any $J \in \mathcal{J}_{m,p}$ the columns of \mathbf{E}_J are an orthonormal basis of \mathbf{R}^J , it follows that $\sum_{i=1}^m \cos^2 \phi_{J,i} = \text{tr}(\widetilde{W}^T \mathbf{E}_J (\widetilde{W}^T \mathbf{E}_J)^T) = \text{tr}(\widetilde{W}^T \mathbf{E}_J \mathbf{E}_J^T \widetilde{W})$. Thus, for some J , $\sum_{i=1}^m \cos^2 \phi_{J,i} \geq \frac{m^2}{p}$, and the inequality is strict for some J unless it is an equality for all J . But for any given J ,

$$\sum_{i=1}^m \cos^2 \phi_{J,i} \geq \frac{m^2}{p} \iff \sum_{i=1}^m \sin^2 \phi_{J,i} = m - \sum_{i=1}^m \cos^2 \phi_{J,i} \leq m - \frac{m^2}{p} = m(1 - \frac{m}{p}), \quad (7.6)$$

and the first inequality in (7.6) is strict if and only if the second is strict. Thus (7.5) holds for some J , and the inequality in (7.5) is strict for some J unless it

is an equality for all J . \blacksquare

Proof of Proposition 5.6.

If $m = p$ then p is even, $R = -I$, and for $\sigma = (-1, -1, \dots, -1)$ we have $I_\sigma = -I$ and $d_{SO}(RI_\sigma, I) = 0 < d_{SO}(R, I)$. Henceforth we assume $m < p$.

Let $W = E_{-1}(R)$ and let $m = \dim(W)$. Note that

$$d_{SO}(R, I)^2 = \frac{m}{2} \pi^2. \quad (7.7)$$

Let $J \in \mathcal{J}_{m,p}$ be such that $\sum_{i=1}^m \sin^2 \phi_{J,i} = \min_{K \in \mathcal{J}_{m,p}} \{\sum_{i=1}^m \sin^2 \phi_{K,i}\}$. By Corollary 7.2, inequality (7.5) holds, and the inequality is strict unless

$$\sum_{i=1}^m \sin^2 \phi_{K,i} = m(1 - \frac{m}{p}) \quad (7.8)$$

for all $K \in \mathcal{J}_{m,p}$. Let $\sigma = \sigma^J$. By Corollary 6.4,

$$d_{SO}(RI_\sigma)^2 = 4d_{Gr}(W, \mathbf{R}^J)^2 = 4 \sum_{i=1}^m (\phi_{J,i})^2 \quad (7.9)$$

where $\phi_{J,i} = \phi_{J,i}(W)$.

The function $f : x \mapsto \frac{\sin x}{x}$ is strictly decreasing on the interval $(0, \frac{\pi}{2}]$. Hence for all $x \in (0, \frac{\pi}{2}]$ we have $\frac{\sin x}{x} \geq f(\frac{\pi}{2}) = \frac{2}{\pi}$, with equality only if $x = \frac{\pi}{2}$; thus for $x \in [0, \frac{\pi}{2}]$ we have $x \leq \frac{\pi}{2} \sin x$, with equality only if $x = 0$ or $x = \frac{\pi}{2}$. Hence

$$d_{SO}(RI_\sigma)^2 = 4 \sum_{i=1}^m (\phi_{J,i})^2 \leq \pi^2 \sum_{i=1}^m \sin^2 \phi_{J,i} \quad (7.10)$$

$$\leq m(1 - \frac{m}{p}) \pi^2 \quad (7.11)$$

$$\begin{aligned} &\leq \frac{m}{2} \pi^2 \quad (\text{since } \frac{m}{p} \geq \frac{1}{2}) \quad (7.12) \\ &= d_{SO}(R, I)^2. \end{aligned}$$

Hence $d_{SO}(RI_\sigma) \leq d_{SO}(R, I)$, and this inequality is strict if any of the inequalities (7.10), (7.11), (7.12) is strict. Inequality (7.10) is strict if $0 < \phi_{J,i} < \frac{\pi}{2}$ for some i , and, by our choice of J , (7.11) is strict unless equality holds in (7.8) for all $K \in \mathcal{J}_{m,p}$.

We claim that at least one of the inequalities (7.10), (7.11) is strict. Assume this is not so. Then, since equality holds in (7.10) with J replaced by any $K \in \mathcal{J}_{m,p}$, it follows that for all $K \in \mathcal{J}_{m,p}$ and $i \in \{1, \dots, m\}$ the angle $\phi_{K,i}$ is either 0 or $\pi/2$, and that $\sum_{i=1}^m \sin^2 \phi_{K,i} = m(1 - \frac{m}{p})$ for all K . But for any $V \in \text{Gr}_m(\mathbf{R}^p)$, there always exists $K \in \mathcal{J}_{m,p}$ for which none of the principal angles $\phi_{K,i}(V, \mathbf{R}^K)$ is $\pi/2$. Choosing such K for our m -plane W , all of the principal angles $\phi_{K,i}$ must therefore be 0 (since they are all either 0 or $\pi/2$). But then $\sum_{i=1}^m \sin^2 \phi_{K,i} = 0 < m(1 - \frac{m}{p})$, a contradiction.

Thus at least one of the inequalities (7.10), (7.11) is strict, so $d_{SO}(RI_{\sigma}) < d_{SO}(R, I)$. ■

We will establish Proposition 5.8 (a weak version of Conjecture 5.7) as a consequence of a different weakened version of Conjecture 5.7:

Proposition 7.3. *Let $m \geq 2$ be even, let $R \in SO(p)$ be an involution of level m , and let $\sigma \in \mathcal{I}_p^+$. If $d_{SO}(RI_{\sigma}, I) < d_{SO}(R, I)$, then $\text{level}(\sigma) < 2m$. (Hence if R is sign-change reducible, then it is reducible by a sign-change of level less than $2m$.)*

This proposition, which we will prove this shortly, reduces Proposition 5.8 into a triviality:

Proof of Proposition 5.8, assuming Proposition 7.3: The only positive even integer less than 2×2 is 2. ■

Proof of Proposition 7.3. Let $m_{\sigma} = \text{level}(\sigma)$. Define l_{\pm} as in Lemma 6.2. From (6.19),

$$\frac{1}{2}(l_+ + l_-)\pi^2 \leq d_{SO}(RI_{\sigma}, I)^2 < d_{SO}(R, I)^2 = \frac{1}{2}m_R\pi^2,$$

so

$$l_+ + l_- < m_R. \quad (7.13)$$

But by (6.5) we have $l_- = l_+ + m_{\sigma} - m_R$, so substituting into (7.13), we have $2l_+ + m_{\sigma} - m_R < m_R$; equivalently,

$$2l_+ < 2m_R - m_{\sigma}.$$

Since $l_+ \geq 0$, we must have $m_{\sigma} < 2m_R$. ■

8. Proofs of Sign-change reducibility results, part II: Proposition 4.14

As noted in Section 5.2, Proposition 5.6 proves part (a) of Proposition 4.14. Thus it remains only to prove part (b) of this Proposition.

The combination of Proposition 5.8 and Proposition 5.11 is what will guide our proof of part (b). To establish the result, it suffices to prove that for $p \geq 11$, the answer to Question 5.5 is no—i.e. that there exist involutions in $SO(p)$ that are not sign-change reducible. Hence it suffices to prove that there exist such involutions of level 2. By Proposition 5.8, it therefore suffices to establish (for $p \geq 11$) the existence of involutions that are not sign-change reducible by a sign-change of level 2; thus it suffices to show that Statement 2 of Proposition 5.11 is false when $p \geq 11$ and $m = 2$. For this, we need only produce planes

in \mathbf{R}^p for which we can show that (5.15) is false for all $J \in \mathcal{J}_{2,p}$. Towards this end, we examine two (families) of examples in which $m = 2$ and $p \geq 4$.

Example 8.1. Let $p = 2k$ or $2k + 1$, where $k \geq 2$. Define vectors $\hat{v}, \hat{w} \in \mathbf{R}^p$ by

$$\hat{v} = \frac{1}{\sqrt{k}} \sum_{i=1}^k \mathbf{e}_i, \quad \hat{w} = \frac{1}{\sqrt{k}} \sum_{i=k+1}^{2k} \mathbf{e}_i.$$

The set $\{\hat{v}, \hat{w}\}$ is orthonormal. Let $W_p = \text{span}\{\hat{v}, \hat{w}\}$, a 2-plane in \mathbf{R}^p . We will compute the principal angles between W_p and \mathbf{R}^J for all $J \in \mathcal{J}_{2,p}$. Write $J = \{i, j\}$, where $1 \leq i < j \leq p$. Let \widetilde{W} be the $p \times 2$ matrix whose first column is \hat{v} and whose second column is \hat{w} . Since the columns of \widetilde{W} are an orthonormal basis of W_p , the principal angles between W_p and \mathbf{R}^J are the arc-cosines of the singular values of $\widetilde{W}^T \mathbf{E}_J$.

First suppose that p is even. We divide the elements $\{i, j\} \in \mathcal{J}_{2,p}$ into two cases: Case I = $\{\{i, j\} : i < j \leq k \text{ or } k < i < j\}$; Case II = $\{\{i, j\} : i \leq k < j\}$. The principal values of the 2×2 matrix $\widetilde{W}^T \mathbf{E}_J$ are easily computed to be 0 and $\frac{4}{p}$ in Case I, and $\frac{2}{p}$ (with multiplicity 2) in Case II. Hence the principal angles are

$$\begin{aligned} \phi_{J,1} &= \frac{\pi}{2}, \quad \phi_{J,2} = \cos^{-1} \sqrt{4/p} \quad \text{in Case I,} \\ \phi_{J,1} &= \phi_{J,2} = \cos^{-1} \sqrt{2/p} \quad \text{in Case II,} \end{aligned}$$

so

$$\min_{J \in \mathcal{J}_{2,p}} \{d_{Gr}(W_p, \mathbf{R}^J)^2\} = \min \left\{ \left(\frac{\pi}{2}\right)^2 + \left(\cos^{-1} \sqrt{4/p}\right)^2, 2 \left(\cos^{-1} \sqrt{2/p}\right)^2 \right\}. \quad (8.1)$$

We will return to (8.1) shortly, but first let us do the analogous computation for p odd. For $p = 2k + 1$, we divide the computation into three cases: Case I = $\{\{i, j\} : i < j \leq k \text{ or } k < i < j \leq 2k\}$; Case II = $\{\{i, j\} : i \leq k < j \leq 2k\}$; and Case III = $\{\{i, j\} : i \leq 2k, j = 2k + 1\}$. The principal values of the matrix $\widetilde{W}^T \mathbf{E}_J$ are 0 and $\frac{2}{p} + \frac{2}{p-1}$ in Case I, $\frac{2}{p}$ and $\frac{2}{p-1}$ in Case II, and $\frac{2}{p-1}$ in Case III. Hence the principal angles are

$$\begin{aligned} \phi_{J,1} &= \frac{\pi}{2}, \quad \phi_{J,2} = \cos^{-1} \sqrt{4/(p-1)} \quad \text{in Case I,} \\ \phi_{J,1} &= \phi_{J,2} = \cos^{-1} \sqrt{2/(p-1)} \quad \text{in Case II,} \\ \phi_{J,1} &= \frac{\pi}{2}, \quad \phi_{J,2} = \cos^{-1} \sqrt{2/(p-1)} \quad \text{in Case III.} \end{aligned}$$

Clearly $\phi_{J,1}^2 + \phi_{J,2}^2$ is larger in Case III than in Case II, so

$$\min_{J \in \mathcal{J}_{2,p}} \{d_{Gr}(W_p, \mathbf{R}^J)^2\} = \min \left\{ \left(\frac{\pi}{2}\right)^2 + \left(\cos^{-1} \sqrt{\frac{4}{p-1}}\right)^2, 2 \left(\cos^{-1} \sqrt{\frac{2}{p-1}}\right)^2 \right\}. \quad (8.2)$$

It follows from (8.1) and (8.2) that

$$\begin{aligned} \lim_{p \rightarrow \infty} \min_{J \in \mathcal{J}_{2,p}} \{d_{Gr}(W_p, \mathbf{R}^J)^2\} &= 2 \left(\frac{\pi}{2}\right)^2 = \frac{\pi^2}{2} \\ &\not\leq \frac{\pi^2}{4} = \frac{m\pi^2}{8} \end{aligned} \quad (8.3)$$

since $m = 2$ in Example 8.1. Hence for large enough p , Statement 2 in Proposition 5.11 is false, and therefore so is Statement 1. This already shows that for all p sufficiently large, there exist geodesically antipodal pairs (U, V) in $SO(p) \times SO(p)$ that are not sign-change reducible. However, to get the quantitative statement in Proposition 4.14(b), we have to continue working.

It can be shown⁴ that for $0 < x \leq 1$,

$$(\pi/2)^2 + (\cos^{-1} x)^2 > 2(\cos^{-1} \frac{x}{\sqrt{2}})^2, \quad (8.4)$$

hence that in (8.1) in (8.2), the second of the two expressions being compared is the smaller. Thus

$$\min_{J \in \mathcal{J}_{2,p}} \{d_{Gr}(W_p, \mathbf{R}^J)\} = \sqrt{2} \cos^{-1}(c_p), \quad \text{where } c_p = \begin{cases} \sqrt{2/p}, & p \text{ even,} \\ \sqrt{2/(p-1)}, & p \text{ odd.} \end{cases} \quad (8.5)$$

Since $m = 2$ in Example 8.1, $\sqrt{m\pi^2/8} = \frac{\pi}{2}$, so equation (8.5) shows that (5.15) (with $W = W_p$) is false for all $J \in \mathcal{J}_{2,p}$ if $\sqrt{2} \cos^{-1}(c_p) \geq \frac{\pi}{2}$; equivalently, if $c_p \leq \cos \frac{\pi}{2\sqrt{2}} \approx 0.4440$. This translates to $2\lfloor \frac{p}{2} \rfloor \geq 2 \sec^2 \frac{\pi}{2\sqrt{2}} \approx 10.14$. Hence the answer to Question 5.5 is definitely “no” for all $p \geq 12$. To complete the proof of Proposition 4.14(b), it remains only to show that this “12” can be reduced to “11”. We will accomplish this with the next example.

Example 8.2. Let $p = 2k + 1$, where $k \geq 2$. Define vectors $v, w, \hat{v}, \hat{w} \in \mathbf{R}^p$ by

$$\begin{aligned} v &= \sum_{i=1}^p \mathbf{e}_i, & w &= \sum_{i=1}^k \mathbf{e}_i - \sum_{i=k+1}^{2k} \mathbf{e}_i \\ \hat{v} &= \frac{v}{\|v\|} = \frac{1}{\sqrt{p}} v, & \hat{w} &= \frac{w}{\|w\|} = \frac{w}{\sqrt{p-1}}. \end{aligned}$$

⁴The authors did not find this exercise in Calculus 1 entirely trivial, but are nonetheless leaving it to the reader.

As in the previous example, $\{\hat{v}, \hat{w}\}$ is an orthonormal basis of a plane W'_p . Just as in Example 8.1, we can compute the principal angles between W'_p and \mathbf{R}^J for all $J \in \mathcal{J}_{2,p}$. We define Cases I and II and III just as in the odd- p case of the previous example. The principal values of the relevant 2×2 matrices are 0 and $\frac{2}{p} + 2/(p-1)$ in Case I, $\frac{2}{p}$ and $\frac{2}{p-1}$ in Case II, and

$$\lambda_{\pm}(p) := \frac{1}{p} + \frac{1}{2(p-1)} \pm \sqrt{\frac{1}{p^2} + \frac{1}{4(p-1)^2}}$$

in Case III. Hence

$$\begin{aligned} \min_{J \in \mathcal{J}_{2,p}} \{d_{Gr}(W'_p, \mathbf{R}^J)^2\} &= \min \left\{ \left(\frac{\pi}{2} \right)^2 + \left(\cos^{-1} \sqrt{\frac{2}{p} + \frac{2}{p-1}} \right)^2, \right. \\ &\quad \left(\cos^{-1} \sqrt{2/p} \right)^2 + \left(\cos^{-1} \sqrt{2/(p-1)} \right)^2, \\ &\quad \left. \left(\cos^{-1}(\sqrt{\lambda_+(p)}) \right)^2 + \left(\cos^{-1}(\sqrt{\lambda_-(p)}) \right)^2 \right\} \end{aligned} \quad (8.6)$$

Numerically, we find that for $p = 11$, the middle line of (8.6) is the smallest of the three lines, so

$$\begin{aligned} \min_{J \in \mathcal{J}_{2,11}} \{d_{Gr}(W'_{11}, \mathbf{R}^J)^2\} &= \left(\cos^{-1} \sqrt{2/11} \right)^2 + \left(\cos^{-1} \sqrt{2/10} \right)^2 \\ &\approx 1.0146 \frac{\pi^2}{4}. \end{aligned} \quad (8.7)$$

Since this number is larger than $\frac{\pi^2}{4}$, the answer to Question 5.5 is no for $p = 11$. This completes the proof of Proposition 4.14. \blacksquare

Remarks 8.3. (1) We considered Example 8.2 only for odd p because for even p , the principal angles $\phi_{J,i}(W'_p)$ turn out to be the same as for $\phi_{J,i}(W_p)$ in Example 8.1. In Example 8.2, we can also compute numerically that for $p = 5, 7$, and 9, we have $\min_{J \in \mathcal{J}_{2,p}} \{d_{Gr}(W'_p, \mathbf{R}^J)^2\} < \frac{\pi^2}{4}$. However, we cannot conclude that the answer to Question 5.5 is “yes” for $p \leq 10$, since we have not proven that this example represents the worst case, i.e. that $\min_{J \in \mathcal{J}_{2,p}} \{d_{Gr}(W'_p, \mathbf{R}^J)\} \geq \min_{J \in \mathcal{J}_{2,p}} \{d_{Gr}(W, \mathbf{R}^J)\}$ for all $W \in \text{Gr}_m(\mathbf{R}^p)$. Thus Question 5.5 remains open for $5 \leq p \leq 10$. However, based on computations, it seems likely to the authors that the largest p for which the answer to Question 5.5 is yes is likely to be closer to 10 than to 4.

(2) The number $\frac{\pi^2}{2}$ in (8.3) is exactly the squared diameter of $\text{Gr}_2(\mathbf{R}^p)$ for all $p \geq 4$. Thus, (8.3) shows that as $p \rightarrow \infty$, the distance between W_p and the *closest* coordinate plane(s) \mathbf{R}^J is approaching the *largest* possible distance between two points in $\text{Gr}_2(\mathbf{R}^p)$.

References

References

- [1] R. Bhattacharya, L. Lin, Omnibus CLTs for Fréchet means and nonparametric inference on non-Euclidean spaces, *Proc. Amer. Math. Soc.* 145 (2017) 413–428.
- [2] L. J. Billera, S. P. Holmes, K. Vogtmann, Geometry of the space of phylogenetic trees, *Adv. in Appl. Math.* 27 (4) (2001) 733–767.
- [3] J. Cheeger, D. G. Ebin, *Comparison Theorems in Riemannian Geometry*, North Holland/American Elsevier, Amsterdam, 1975.
- [4] J. Damon, J. Marron, Backwards principal component analysis and principal nested relations, *J. Math. Imaging and Vision* 50 (1) (2014), 107–114.
- [5] A. Edelman, T. A. Arias, S. T. Smith, The geometry of algorithms with orthogonality constraints, *SIAM J. Matrix Anal. Appl.* 20 (2) (1998) 303–353.
- [6] C. G. Gibson, K. Wirthmüller, A. A. du Plessis, E. J. N. Looijenga, *Topological Stability of Smooth Mappings*, Lecture Notes in Mathematics, Vol. 552, Springer-Verlag, Berlin, 1976.
- [7] G. H. Golub, C. F. Van Loan, *Matrix Computations*, 2nd edition, The Johns Hopkins University Press, 1989.
- [8] D. Groisser, S. Jung, A. Schwartzman, Geometric foundations for scaling-rotation statistics on symmetric positive-definite matrices: minimal smooth scaling-rotation curves in low dimensions, preprint (2017).
URL <https://arxiv.org/abs/1602.01187>
- [9] D. Groisser, S. Jung, A. Schwartzman, A scaling-rotation metric on the space of symmetric positive-definite matrices, in preparation.
- [10] T. Hotz, S. Huckemann, H. Le, J. S. Marron, J. C. Mattingly, E. Miller, J. Nolen, M. Owen, V. Patrangenaru, S. Skwerer, Sticky central limit theorems on open books, *Ann. Appl. Prob.* 23 (6) (2013) 2238–2258.
- [11] S. Jung, A. Schwartzman, D. Groisser, Scaling-rotation distance and interpolation of symmetric positive-definite matrices, *Siam J. Matrix Anal. Appl.*, 36 (3) (2015) 1180–1201.
- [12] D. G. Kendall, D. Barden, T. K. Carne, H. Le, *Shape and Shape Theory*, Wiley Series in Probability and Statistics, John Wiley & Sons Ltd., Chichester, 1999.
- [13] A. Schwartzman, Random ellipsoids and false discovery rates: statistics for diffusion tensor imaging data, Ph.D. thesis, Stanford University (2006).

- [14] A. Schwartzman, W. F. Mascarenhas, J. E. Taylor, Inference for eigenvalues and eigenvectors of Gaussian symmetric matrices, *Ann. Statist.* 36 (6) (2008) 2886–2919.
- [15] Y.-C. Wong, Differential geometry of Grassmann manifolds, *Proc. Nat. Acad. Sci. U.S.A.* 57 (1967) 589–594.
- [16] H. Zhu, H. Zhang, J. G. Ibrahim, B. S. Peterson, Statistical analysis of diffusion tensors in diffusion-weighted magnetic resonance imaging data, *J. Amer. Statist. Assoc.* 102 (480) (2007) 1085–1102.